DOCUMENT RESUME

ED 066 967 FL 003 375

AUTHOR Feinstein, Stephen H.; And Others
TITLE Communication Sciences Laboratory Quarterly Progress Report, Volume 9, Number 3: Research Programs of Some of the Newer Members of CSL.
INSTITUTION Florida Univ., Gainesville. Communication Sciences Lab.
PUB DATE 71
NOTE 78p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Acoustic Phonetics; Acoustics; Artificial Speech; *Audition (Physiology); Auditory Perception; Aural Stimuli; Blind; Communications; Deaf Research; *Echolocation; *Information Systems; Intonation; *Language Research; Nonverbal Communication; Oral Communication; Physiology; Speech; Tone Languages

ABSTRACT

The research reported in these papers covers a variety of communication problems. The first paper covers research on sound navigation by the blind and involves echo perception research and relevant aspects of underwater sound localization. The second paper describes a research program in acoustic phonetics and concerns such related issues as consonant-vowel transitions in the speech of deaf adults and the intelligibility of whispered speech in a tone language. The third paper studies some of the basic problems in data-sharing among information systems and in ambiguity resolution and feedback utilization in man-machine communication. The final paper deals with studies in psychoacoustics and describes research in binaural hearing and two sets of interactions involving the interaural differences required to localize a sound source. (VM)

ED 066967

COMMUNICATION SCIENCES LABORATORY

QUARTERLY PROGRESS REPORT

9.3, FALL, 1971

Research Programs of Some of the Newer Members of CSL

Preface

The purposes of these progress reports are:

1. to provide other laboratory investigators and professional workers in the field with up-to-date information about our research activities and results,
2. to serve as documentation of our research activities for agencies which provide us with support,
3. to provide somewhat formal reporting of research activity for our own faculty and students in order to exchange information and encourage collaborative efforts.

In this report we have chosen to present material from some of our younger faculty. This, it is hoped, will acquaint our readers with their contributions to our research program and will demonstrate that we have been significantly enriched by them.

Inquiries concerning these reports should be addressed to the editor, Robert J. Scholes.

Table of Contents

Stephen H. Feinstein

Dr. Feinstein was educated at Florida State University (BS, 1960; MS, 1962) and Dalhousie University in Halifax, Nova Scotia (Ph.D., 1971). His major areas of study were psychology and biology.

Dr. Feinstein joined the CSL group as a post-doctoral fellow in 1970.

Howard B. Rothman

Dr. Rothman was educated at the City College of New York (BA, 1961), New York University (MA, 1964), and Stanford University (Ph.D., 1971) majoring in music, speech pathology, and speech and hearing sciences at the respective degree levels.

He joined CSL as a research instructor in 1969.

His major research interests are acoustic phonetics and underwater communications.

Stanley Y. W. Su

Dr. Su was educated at the Tamkang College of Arts and Sciences (BA, 1961), the University of Wisconsin (MS, 1965; PH.D., 1968). His major fields were Languages, Mathematics, and computer sciences.

He has had professional experience as a computer programmer and mathematician at the University of Wisconsin and the RAND corporation.

Stanley joined CSL as a joint appointee with Electrical Engineering in 1970.

William A. Yost

Dr. Yost received his degrees from The Colorado College (BA, 1966) and Indiana University (Ph.D., 1970) majoring psychology.

In addition to the usual research and teaching assistantships, Bill was a post-doctoral fellow in the psychology department at UC, San Diego, in 1970-71.

He joined our group this fall (1971), taking over the program in psychoacoustics formerly run by Dr. John Brandt.

Stephen H. Feinstein

I. Personal Research

1. Research on Sound Navigation by the Blind

a. Introduction

The location and avoidance of objects by blind people has been the subject of scientific inquiry for over 200 years, beginning with Spallanzani's experiments with bats (Dijkgraaf, 1960). Many other organisms are now either known, or suspected of, making use of acoustic signals for navigation and orientation in space. However, it was not recognized, at first, that the perception of objects by the blind is the same in principle as the echo-ranging of the bat and the porpoise. The skin of the face of the blind person was presumed to be especially sensitive to air currents which were detected as an object was approached (Diderot, 1961); the name given this phenomenon was facial vision. Taylor (1966, p. 80) suggests that the cutaneous sensation reported by blind people as they approach an object is the result of motor and labeling responses being conditioned to the same set of stimuli.

It is now known that the mechanism used to detect objects is actually sonar (SOund NAvigation and Ranging), and that it depends entirely upon sonic echoes (Ammons, Worchel, and Dallenbach, 1953; Cotzin and Dallenbach, 1950; Supa, Cotzin, and Dallenbach, 1944). The sonar response occurs as a function of either the sonic environment generated by the listener (active sonar) or the ambient sonic environment (passive sonar).

All of the early research on echo perception dealt with how well blind people could avoid certain barriers and obstructions, or with the demonstration of the nature of the response. Experiments by Kellogg (1962) indicated that it was possible to apply psychophysical methods to the echo-ranging process as it is used by the blind. The results of these experiments are only an approximation of the actual acuity of the system in terms of size and distance perception. However, this position can be ascribed to the naivete of the experimenter who assumed that special modifications in the auditory system were necessary for any greater acuity (Kellogg, personal communication) and so did not attempt to use very small targets or very small differences in size and distance. One surprising result of this experiment was that at a distance of one foot, differences in surface density of various materials (like velvet and denim) could be detected with great accuracy. As a direct result of Kellogg's (1962) work, a series of experiments was undertaken to more accurately determine the acuity and extent of human echo perception (Rice, 1967; Rice and Feinstein, 1965a, 1965b; Rice, Feinstein and Schusterman, 1965). These experiments indicate that: (1) For distances between 2 and 9 feet, the absolute threshold for detection of circular metal targets is approximately $4\frac{1}{2}^{o}$ of radial angle (about $1\frac{1}{2}$ inches in diameter at 2 feet); (2) For distances between 2 and 4 feet, difference thresholds are approximately 1.11/1 (.20 inch difference in diameter at 3 feet); (3) At a distance of 67 inches, subjects are able to localize a target in 180^{o} of azimuth with a standard deviation from center of 10.1^{o} when the target subtends an angle of 10.2^{o}; (4) Target detection is possible with little loss of acuity when one ear is blocked; and

(5) Simple shape discrimination (circle, square, and triangle with the same surface area) is possible with approximately 80% accuracy at a distance of 3 feet.

Several investigators have reported instances of sighted subjects learning to avoid obstacles in less than 100 trials (Supa, Cotzin, and Dallenbach, 1944; Worchel and Ammons, 1945) or learning to perform fairly difficult size, size difference, and shape discriminations under controlled laboratory conditions (Rice, 1967). Acquisition of the response has generally been of secondary interest but two experiments which bear on this problem are at hand. Worchel and Mauney (1951) worked with seven totally blind subjects who had failed an obstacle detection test by making two or more false alarms or more than five collisions in 54 trials for a total of 210 obstacle exposures; two series with a five-minute rest between were given each day. The experiments used the following procedures "in order to facilitate learning as much as possible:" (1) Punishment -- collisions with obstacles were not prevented; (2) Reward -- a pat on the back and praise if detection was made one foot or less from the obstacle; (3) Withholding reward -- if performance was poor or the appraisal was made at more than one foot; (4) Knowledge of Results -- after every appraisal the subject was led up to the target so he would know the amount of his error. After completion of the training trials, a 60-trial test with no feedback was given with the result that all seven subjects showed greater consistence in "first perceptions" regardless of the starting point from the obstacle, smaller and more consistent final appraisals, and fewer collisions.

Taylor (1962) has proposed that all conscious experience is the result of an acquired and specific readiness to respond to objects in the environment. The organism in its environment, as a kind of multistable system, is capable of coping with disruptions in sensory input by adapting the necessary subsystem(s). The process of adaptation occurs in two stages: (1) The subsystem randomly adopts different values of each of its variables, until a combination occurs which overcomes the disruption; (2) Neural connections or engrams are established as a result of conditioning. A proper test of this hypothesis (Taylor, 1966, p. 64) would be to take "...a class of stimuli that do not ordinarily give rise to any perceptual experience and condition them to a class of responses that they have not previously evoked." The procedure used by Taylor (1966) was extremely simple. The experiment was conducted in large reverberant room (no ambient noise level is given) which contained, besides the experimenter and the subject, a table measuring 4 x 8 feet, a chair, and a target mounted on a small stand. The subject, wearing a blindfold, sat at the middle of one of the long sides of the table. The opposite side of the table was marked off into eight equal divisions which could be used to guide the experimenter's placement of the targets. Appropriate measures were taken to avoid giving the subject any extraneous information about the movement of the targets. The subject was instructed to "...scan the field by rotating the head about its vertical axis while repeatedly uttering words such as 'where is it?' and to stretch out a hand whenever he thought there was a possibility that he might have received a signal (sic) in the direction from which it might be coming. If he touched the target the trial would end, but if he failed he was to withdraw his hand and continue the search. He was not allowed to search for the target with a sweeping movement of the hand." Ten trials were given and then the experimenter questioned the subject about his subjective experience and the extent to which he had to guess. The question period was followed by an additional 10 trials (some subjects had as many as 300 trials while others had no more than 50). Taylor makes much of the assertion by two of his subjects that they did not associate any sensation with their perception of the targets. It is the present writer's personal experience that none of the subjects he has tested (15 in all) ever had the slightest doubt about the cue to which they were attending although there was some variability in the labels which they attached to those cues.

Thus, echo perception research began with qualitative description and a few isolated attempts to manipulate acquisition; then quantitative measurements of a psychophysical nature were undertaken. All of the experiments have been designed within the framework of the active sonar paradigm. That is, the dependent variable has been some function of target detection, viz, number of collisions or detections, error in location of targets, or number of correct identifications of targets of various shapes and materials. While these experiments provide very useful information about the potential sensitivity of the human active sonar mode, they do not assess the potential of the passive mode. The former is particularly useful to the listener when nearby objects are to be located relative to his position but it is of little value as a device for navigation. The latter provides some information about the location of objects but is more suited to the problem of navigation over relatively long distance For example, the blind listener may use the sound of traffic to guide him to a road and the clicking of a traffic light relay box to bring him to a particular corner (passive sonar). On his way from the starting point to the corner he will avoid lamp posts, baby carriages and other obstacles by generating sounds with his cane or some other source, and listening for the echoes which these objects create (active sonar).

It is clear that at this point we know a great deal about a portion of the human sonar response; however, the known part represents only about one-third of the problem. That is, we know about the active mode but we have no quantitative data on the passive mode, nor do we have hard data relating the total response to potential environmental design for the sightless population. The aim of the proposed research program is to provide data concerning these issues.

b. Procedures

(1) Passive Sonar Performance

These experiments are designed to provide information about the functional limits of the passive sonar response. That is: 1) How accurately can a listener locate a distant sound source in terms of azimuth and distance? 2) Can the listener navigate to the source efficiently? and 3) What are the acoustic parameters which determine the effectiveness of the passive sonar response?

All of the proposed experiments in this portion of the program will be carried out in a large, open and relatively flat field. This field (already selected) is sufficiently large to provide a navigation range approximately 200 yards in diameter. Although it is located on the campus of the university, ambient noise levels are low enough for purposes of the research. The subject (S) stands at the center of the circular range (marked in 10^{o} intervals) and the sound sources are moved to predetermined points on the perimeters of a series of ten concentric circles. Thus, on any trial the sound can be located at any one of 360 positions.

(a) Stationary Localization of a Sound Source

In this experiment the S (blind) will remain at the center of the course. His task will be to make a verbal estimate of the distance of the sound (in feet) and to point at the sound with a light rod (3 ft.) which suspends a plumb line. The experimenter (E) will mark the point at which the plumb line rests and will later draw a line from the center of the circle through the mark to determine the pointing error. These data will provide an estimate of the precision with which S can localize distant sound sources.

(b) Passive Sound Navigation

The second experiment will determine the efficiency with which a sightless S can navigate to a sound source. Seven different sounds will be used, viz, 250 Hz, 500 Hz, 1000 Hz, sinusoid or pulsed and pulsed broadband white noise. The subject will walk from the center of the circle to the source as quickly as possible. A small lightweight device which will leave a thin white line on the ground will be pulled by S as he walks. At the end of each run, E will stretch a string from the center of the circle to the sound source and then measure the distance of the white line to the string at 2 yard intervals. These data will provide a measure of the efficiency of sound navigation.

(c) Acoustic Parameters

In the final experiment, the navigation experiment will be replicated in the presence of running and/or stationary motor vehicles, groups of pedestrians, and large objects which mimic the walls of buildings. An attempt will be made to determine which signals are most effective and what happens to the efficiency of the response in simulated traffic.

(2) A Sonic Environment for the Blind

The next stage in the research program will be to design a model beacon and reflector system based on all of the data collected over the past two decades. We will have available to us sufficient psycho-physical data to select the most effective combination of materials, signals, and locations. The proposed site for this pilot program is the Florida School for the Deaf and Blind at St. Augustine, Florida. This institution is one of the largest of its kind in the country and would provide the opportunity for testing of the system on a relatively large population.

c. Significance

Obviously, the immediate benefit to be derived by the blind members of the community makes this work both relevant and significant. This research is also critical to the program suggested by Hollien and Feinstein for the investigation of the sound localization response to be used by divers under conditions of poor visibility (see next section). In that instance, it is imperative that we obtain a performance baseline in air with which we can compare performance in the sea.

References

Ammons, C. H., Worchel, P., and Dallenbach, K. M., "Facial vision": The perception of obstacles out-of-doors by blindfolded and blindfolded deafened subjects, Amer. J. Psychol., 66, 519-553, 1953.

Cotzin, M., and Dallenbach, K. M., "Facial vision": The role of pitch and loudness in the perception of obstacles by the blind, Amer. J. Psychol., 63, 485-515, 1950.

Diderot, D., Letter on the blind, 1749, in Early Philosophical Works, Jourdain, M. (Trans.), Chicago: Open Court, 68-141, 1961.

Dijkgraaf, S., Spallanzani's unpublished experiments on the sensory basis of object perception in bats. Isis, 51, 9-20, 1960.

Kellogg, W. N., Sonar system of the blind, Science, 137, 399-404, 1962.

Kellogg, W. N., Auditory scanning in the dolphin, Psychol. Rec., 10, 25-27, 1960.

Kohler, I., Echo perception in the blind, Amer. Found. for the Blind Res. Bull., 4, 1-4, 1964.

Rice, C. E., Human echo perception, Science, 155, 656-664, 1967.

Rice, C. E., and Feinstein, S. H., The influence of target parameters on a human echo detection task, Proc. Amer. Psychol. Assn., 65-66, 1965.

Rice, C. E., Feinstein, S. H., and Schusterman, R. J., Echo-detection ability of the blind: size distance factors, J. Exp. Psychol., 70, 246-251, 1965.

Supra, M., Cotzin, M., and Dallenbach, K. M., "Facial vision": The perception of obstacles by the blind, Amer. J. Psychol., 57, 133-183, 1944.

Taylor, J. G., Perception generated by training echo location, Can. J. Psychol., 20, 64-81, 1966.

Taylor, J. G., The behavioral basis of perception, New Haven, Conn: Yale Univ. Press, 1962.

Worchel, P., and Berry, J. H., The perception of obstacles by the deaf, J. Exp. Psychol., 43, 187-194, 1952.

Worchel P., and Dallenbach, K. M., "Facial vision": Perception of obstacles by the deaf blind, Amer. J. Psychol., 60, 502-553, 1947.

Worchel, P., and Mauney, J., The effect of practice on the obstacles perception by the blind, J. Exp. Psychol., 41, 170-176, 1951.

II. Interdisciplinary Research

1. Underwater Sound Localization (with Prof. Harry Hollien)

a. Introduction

In order to appreciate the phenomenon of diver sound localization it is first necessary to consider sound localization in air, then to determine the effect of the change in medium on the primary localization cues, and finally to consider the kinds of changes that have taken place in the auditory systems of some marine mammals to enable them to receive sound in water effectively.

By 1930, when Firestone reported measurements of interaural time differences (ITD) and interaural intensity differences (IID) on an artificial head, the primacy of those two stimuli in auditory localization (Molino, 1970) had been firmly established. His data agreed with earlier predictions (Stewart, 1914; Hartley and Fry, 1921) except for minor differences in the obtained IID's caused by interference of the neck not accounted for in the early equations and differences at 1944 Hz where phase cancellation at half-wave length distances altered the diffraction pattern. Later, Nordlund (1962a) replicated Firestone's experiment, but with a more precisely detailed head and greater control over the azimuth and reception of the source. He found that the ITD "...constitutes a linear function of azimuth between the angles of 0^{o} - 60^{o} and between 120^{o} and 180^{o}. The same thing applies to phase differences, if expressed in time. The difference in intensity appears to be an irregular function of azimuth and has to a great extent failed to agree with previously published findings" (p. 75). Nordlund (1962b) also determined that below two kHz the IID was in agreement with earlier predictions and findings (Hartley and Fry, 1921; Firestone, 1930). The ITD were related to azimuth almost exactly as predicted by Woodworth's (1938) theoretical equation derived from the path differences around a rigid sphere. Finally, Fedderson, Sandel, Teas and Jeffress (1957) employed probe microphones inserted into the ears of live listeners and obtained results that agreed with Woodworth's predictions for ITD and approximated the IID's predicted by Hartley and Fry and obtained by Firestone. However, they did not find the intensity versus azimuth irregularity reported by Nordlund.

Another relation is germane to this issue. In 1907, Rayleigh developed equations predicting that interaural phase differences will only operate below a limiting frequency of 1500 Hz and interaural intensity differences will only operate above that frequency. These equations have proved correct; for example, the studies conducted by Stevens and Newman (1934, 1936) are a case in point. These workers had their subjects perched on top of a 12-foot pole mounted on the roof of the laboratory; a small speaker also mounted on a 12-foot pole rotated around them. The experimental signals were sinusoids and subjects responded to them at 13 different positions. The authors found 1) a considerable proportion of front back reversals, 2) localizations accurate below two kHz, 3) poorer localizations from 2 to 4 kHz, and 4) improved localization at higher frequencies. Others have replicated (and confirmed) these findings directly or indirectly (Mills, 1958; Fedderson, et. al., 1957; Sandel, Teas, Fedderson and Jeffress, 1955; Shaxby and Gage, 1932).

In any case, to reasonably summarize relevant (to the underwater milieu) data concerning sound localization in air, certain relationships are generally agreed upon: 1) Precision of localization is greatest with complex sounds such as clicks or thermal noise (Sedee, 1957; Sandel, et. al., 1935; Stevens and Newman, 1934, 1936); 2) For any stimulus, the most precise localization occurs in the median plane (Jeffress and Taylor, 1961; Mills, 1958; Sandel, et. al., 1955; Stevens and Newman, 1934); and 3) Depending on the composition of the stimulus, its intensity, and the psychophysical methods used, the precision of the localization can vary from 1° to 20° (Green and Henning, 1969).

It is common knowledge that the physical characteristics of sound in water are different than in air. To begin with, the auditory apparatus is housed in a flesh covered bony case having an acoustic impedance very closely approximating that of water; this relationship can be given by the equation $Z = \rho c$ where ρ is the density of the medium in grams per cc and c is the velocity in cm per second. Further, the equation yields an acoustic impedance in air and water approximating 41.5 gm/cm^2/sec and 150,000 gm/cm^2/sec. If the Z values of two media are close, the acoustic energy passes easily from one medium to the other. If, however, the differences are large, most of the energy will be reflected and will not pass. It follows then, that in air the head forms an effective acoustic barrier and the pinnea can function as additional sound channeling devices. This means that binaural cues such as interaural intensity differences (IID) and interaural time differences (ITD) as well as possible monaural cues such as the postulated differential reflection of sound waves by the convolutions of the pinnea (Batteau, 1967, 1968) may determine or influence sound localization in air.

In water there is practically no impedance mismatch between the fluid and the head; hence the pressure waves can travel into and through the head with little loss of energy due to reflection and, under these conditions, it is no longer a serious barrier to sound waves. Accordingly, although IID and ITD still exist, they are radically changed from those normally experienced in air. In water, IID must be limited by the energy loss over the linear distance between the cochlea, a distance of approximately 9 cm when measured immediately rostral to the internal auditory meatuses. Moreover, ITD also is limited by the linear distance between the cochlea and in this case the delay for a given distance is 4.5 times less than that for the same distance in air. That is, the velocity of sound in air is described by the equation $c = f\lambda$ where c is the velocity, f is the frequency, and λ is the wave length; at a temperature of 20° C, c = 33,160 cm/sec in air. In water, the velocity of sound is independent of frequency so that for sea water at any temperature (centigrade) (t), salinity (parts/thousand) (S), and depth (D in cm), velocity is determined by the equation (Albers, 1965) $c = 141{,}000 + 421t - 3.7t^2 + 110S + 0.018D$. At a temperature of 10° C, a salinity of 35 ppt, and a depth of 110,488 cm, the velocity of sound is 150,677 cm/sec or 4.5 times faster than the speed of sound in air. Assuming a maximum distance of 9 cm between adult cochlea, the time delay occurring at 90° or 270° would be approximately 60 microsec. (Our calculations as well as Tobias and Zerlin, 1959, and Zerlin, 1970.)

It was assumed that these transformations imposed on the auditory cues in water would necessitate special anatomical adaptations if effective sound reception and localization were to take place. Indeed, in some marine mammals this has been the case. Norris (1964) has summarized the adaptations to water-borne sound to be found in the dolphin (odontocete) ear as follows: "Each inner ear of odontocetes is encased in a very dense, almost ivory-like bone, suspended by ligament and nerve inside a cavity filled with a stable mucus foam...This buble-filled barrier is the best of acoustic insulators, serving to isolate each inner ear from the other and both from the animal itself. The volume and pressure within these barriers can apparently be adjusted to ambient pressure changes as the animal dives...The mechanical sound transmission linkage of the middle ear bones is directly adjusted to the sixty-fold increase in pressure amplitude of a sound wave in water as compared to the same sound wave in air. The external ear canals may be either complete tubes, reduced in part to a ligament, or they may be blocked at the skin surface. The function of these canals remains equivocal."

No such adaptation has occurred in man, a recent and only occasional visitor into the sea. Hence, based on the transformations of the binaural localization cues imposed by the water and on the absence of special anatomical adaptations comparable to those found in the odontocetes, it was assumed that underwater localization would be impossible for man (Bauer and Torick, 1965; Dudok van Heel, 1959; Kitagawa and Shintaku, 1957; Miles, 1966; Reysenback de Haan, 1957; Sivian, 1947). This position received some empirical support both from Bauer and Torick (1965) and from Kitagawa and Shintaku (1957) who produced "high percussion sounds" by hitting two small stones or two small bottles together. These sound sources were then moved back and forth in front of the subjects who reported apparent location. Except for "certain individual variations" and changes caused by head movement, the "sound image" was fixed near the occipital region. When the buzzer was used as a sound source the image appeared to be fixed at a distance of 30 cm in front of the forehead regardless of the position of the source. If the subject closed his ears with his fingers, the image shifted to the occipital region. These authors do not give any information about the physical characteristics of the test situation, e.g., fresh or salt water, pool or lake, distance to source, and so on--so that it is difficult to evaluate their results. Further, Reysenbach de Haan (1957) reported that his test subjects were not able to localize an underwater sound source at "various distances." No details of the experiment are given in this report either; accordingly, it is as difficult to evaluate de Haan's negative results as it is to evaluate Kitagawa and Shintaku's.

In 1944, Ide wrote a classified report in which he described his attempts to utilize the sound localization ability of commando swimmers to home on a target. He found that the sound produced by an ammonia jet could be localized immediately by some divers and by others after a few hours practice with an "anti-masking helmet" (a four-inch strip of foam rubber one-half inch thick running from the forehead over the top of the head to the base of the skull).

Ide's (1944) report was not cited in the underwater localization literature until Feinstein (1966) reported his attempt to determine whether underwater localization could be demonstrated in the laboratory. His localization tests were run in both a reverberant and in an anechoic tank. A boom suspended a projector (underwater loud speaker) in front of the subject and he indicated whether it was to his right or his left. Some divers experienced difficulty in localizing the sound at first but were able to do so eventually; others were able to localize immediately. Feinstein concluded that his subjects could at least determine the quadrant from which the sound came.

Somewhat later, Hollien (1969; 1970; 1971) reported a number of experiments designed to determine if the extent of underwater sound localization could be established and explained. His experiments were conducted using an apparatus (based on DICORS: Hollien and Thompson, 1967) which presented signals from any one of five positions (90^o, 45^o, 0^o, 315^o, 270^o) and the subject's task on any trial was to choose one of these five positions as the location of the signal. In his first experiment, Hollien (1969) found that all 17 of his subjects were localizing at far greater than chance levels (20% correct responses) when the signals were 250 Hz, 1 kHz, 6 kHz and thermal noise presented at 40 dB re hearing threshold. In his next experiment, (Hollien, Lauer and Paul, 1970), he found no essential difference in percent of correct responses when subjects' heads were mobile or immobile but he did find that performance improved with practice. A variety of signals and three intensities were used in the next experiment, (Hollien, Stouffer and Lauer, 1971). The most effective signals were found to be thermal noise, at 50 pulses per second and glides from 100-400 Hz and 2000-500 Hz. The hypotheses presented are tentative and considerable further research is necessary if this facility (in man) is to be understood and developed.

Recent pilot experiments by Feinstein (1969) were conducted in open water at a depth of 30 feet on a Canadian Force's diving tender moored in 65 feet of water; test subjects were professional navy divers. A testing platform (also based on DICORS) constructed of PVC tubing contained a chair in which the diver sat, a head rest to maintain the diver's head in the proper position, and two aluminum "T" beams (1" x $1\frac{1}{2}$" x 8') from which two Chesapeake J-9 projec ors were suspended. Signals were 0.20 sec in duration and came 0.3 sec apart; both sinusoids and thermal noise were used as signals. Subjects signaled when they were ready to listen without breathing. At that point a signal was presented either to the right or left speaker and subjects signaled their decisions via a hand line. Sinusoids (3.5 kHz and 6.0 kHz) were both difficult to hear and hard to locate; thermal noise, however, was easier to hear and could be located with at least 75% accuracy when the projectors were 5^o on either side of the median plane. All but two of the 10 divers tested were able to localize. These experiments are important in that they indicate that heavy arctic wet suits and extremely cold weather, as well as very noisy conditions, do not mitigate the ability to localize sounds.

Leggiere, McGriff, Schenck and van Ryzin (1970) found their subjects could point to a sound source 40 feet away with an overall pointing error of 58 degrees. They used pure continuous sinusoids of .6, .8, 1, 2, 4, and 6 kHz and reported that, "There did seem to be a suggestion, however, that the low frequency (600, 800, 1000 Hz) tests and the high (2000 and 4000 Hz) frequency tests might come from different populations." They did not find pulsed signals to be more effective nor did they find improvement in performance with practice. These investigators also found that their subjects were able to swim 40 feet to a pulsed 800 Hz signal within five minutes on 12 out of 20 trials.

Andersen and Christensen (1969) examined directional hearing in seven divers tested with 1, 2, 4, 8, and 16 kHz one-second pulses. The experiment was carried out at a "free field station" and in a "harbor enclosure." The divers responded "right" or "left" of the median plane when one of 12 sound projectors at ± 10°, 15°, 20°, 30°, 45°, or 90° was energized. These authors concluded that "Directional hearing underwater seems to work on the same parameters as in the air, with allowances for the longer wavelengths in water. At 1 kHz, time/phase cues seem to be effective, and above 4 kc/s directional hearing is supported by intensity differences. Performance improved with increasing frequency from 2-16 kc/s."

Finally, Feinstein's doctoral thesis (1971) was designed to determine the acuity and precision of the sound localization capacity of divers. He found that, with training, acuity for minimum audible angle improved and that performance ranged from 11.5° at 3.5 kHz to 11.3° at 6.5 kHz to 7.5° with broadband thermal noise. Precision, defined as the correspondence between the objective azimuths, was found to be surprisingly good, e.g., for thoroughly trained subjects, the mean deviation for 24 angles was 17.9° for white noise, 26.05° for 3.5 kHz, and 17.8° for 6.5 kHz. Even more interesting was the replication of the finding that the higher frequency sinusoid yielded more precise localization than did the lower one.

b. Objectives

It is now apparent that humans do have some facility for localizing sounds underwater. However, very little is known of the nature and extent of this characteristic--and much relevant information is needed. Accordingly, the objectives of the proposed research are: 1) to determine the stimulus characteristics which control underwater sound localization, and 2) to provide information concerning the mechanism, as well as the extent, of underwater sound localization. The following experiments were designed with these objectives in mind.

(1) Physical Factors Involved in Angular Localization

The perceptual effects of manipulating time, phase, and intensity interaural differences were established in the localization literature first and later the acoustic differences were determined by Nordlund (1961) and others. We will enter the problem of underwater localization in the reverse order, i.e., we will first determine the magnitude of the acoustic differences which exist and then we will manipulate these differences.

Some of the five proposed experiments will replicate those reported by Nordlund (1961) who used an artificial head to determine time, phase, and intensity differences associated with different azimuths. We propose to place hydrophones at the external meatus of our subject rather than to use an artificial head. The reasons for these procedural variations are: a) we are interested in the effect of the whole body on the acoustic signals and b) we could not replicate in an artificial head, and in their proper relationships, all of the air filled spaces and other reflecting surfaces (sinuses, mask, regulator, tanks, etc.) exhibited by a diver. A second set of experiments will deal with the transmission of sound between the water and the cochlea.

In this case, acoustical measurements will be made of the transmission characteristics of an animal (pig) head and if the results warrant, a fresh human cadaver will be used also.

(2) Experiments in Passive Underwater Sound Localization

The second set of experiments will investigate certain of the mechanisms underlying sound localization in water. The applicants and others have obtained data which describes the sensitivity of the submerged human ear to various sound stimuli and the acuity and precision that divers may bring to bear on the problem of sound localization. However, there is still considerable controversy about the specific mechanisms involved in underwater hearing and localization. Hence, the experiments in this section are divided into three categories: a) nonauditory sound localization, b) the relationship of binaural differences to underwater sound localization and c) phantom sound source experiments in which time, phase, and intensity cues are manipulated.

c. Procedure

(1) Physical Factors Involved in Angular Localization

At present, there are no data available which describe objectively the acoustic stimulus as it arrives at the ears of the underwater listener. It only is possible to estimate differences in time of arrival, phase, and intensity based on our hypotheses about the way sound is received at the head. Yet, in order to understand the localization process, we must be able to specify the stimuli. To that end, the following experiments are proposed:

(a) Interaural Acoustic Measurements External to the Head

In all of the experiments to be described in this section, the subject's head will be precisely positioned and maintained by having him don a face mask rigidly attached to DICORS. The sound source and DICORS will be suspended from a positioning system normally used for the calibration of hydrophones and transducers. The source will move around the subject's head at a distance of six feet, rotation will be clockwise and counterclockwise for each trial. Small hydrophones will be mounted against the head and partially covering the meatus.

Experiment 2a: Determination of interaural time differences (ITD) as a function of azimuth. A J-9 projector will emit a train of square waves within an impulse frequency of 360 per second as it circles the subject's head. The output of the hydrophones will pass through amplifiers and into a two channel CRO. The time ratio between the two ears will be read off directly by measuring the horizontal distance between identical points on the two curves. The projector will be moved in 10° steps so that there will be 36 points determined for each frequency. Data will be obtained on at least five subjects.

Experiment 2b: Determination of interaural phase differences (IPD) as a function of azimuth. Nordlund (1961) pointed out that "The interaural phase difference can be expressed in time, in which case it can be regarded as time difference of pure tones. Interaural phase difference, therefore, can be measured in the same manner as the interaural time difference simply by letting the speaker voice be stimulated by pure tones..." In the present experiment, which is a replication of experiment 2a, the stimuli will be 1, 2, 4, 8, and 16 kHz pure tones. The measurements will be completed on five subjects.

Experiment 2c: Determination of interaural intensity differences (IID) as a function of azimuth. The hydrophones will feed a graphic level recorder which will be calibrated with the azimuth of the projector. The projector will be rotated around the subject's head and sound level will be recorded continuously. The dependent variable will be the difference between the curves obtained from the right and left ears. Data will be obtained on five subjects.

(b) Interaural Acoustic Measurements at the Cochlea

Presently there is no information about the way sound is transmitted into the cochlea when the head is submerged. The assumption is that the pressure wave is transmitted via the bones of the skull, i.e., hearing in water is comparable to bone conduction in air. It is suggested that, because of the favorable impedance match between the head and the surrounding water, in this case, transmission of sound may be considered to be from one fluid medium to another. Hence, it follows that under these conditions, there would be an insignificant amount of reflection from the head as the sound wave passed through it and each hearing organ would be activated as the plane wave reached it.

Two experiments are proposed which will attempt to establish the mode of transmission of the sound energy into the cochlea. The first will indicate whether the energy lost between the source and a point behind the head is determined solely by distance and the second will determine whether the interaural time delay (ITD) and interaural intensity difference (IID) are direct functions of the distance between the cochlea and the azimuth of the source. If transmission is via bone conduction there must be energy loss at the head due to reflection and absorption of the pressure wave, and in that case, the first experiment will indicate that energy loss is greater than that predicted as a function of distance alone. If transmission occurs as from one fluid to another (of the same impedance) then the IID and ITD functions should have (equal) slopes determined by the linear distance between the ears and the azimuth of the source.

Four pig heads will be obtained from a local abattoir for use as acoustical experimental material. Each head will be used within the first three hours after death so that there will be as little fluid loss and tissue change as possible. The acoustic signals utilized will be as follows: .1, 1, 3.5, 6.5, 10, 20 kHz and broadband thermal noise. The signals will be pulsed with a duration of less than 10 msec and the rise-fall time will be less than 1.0 msec.

In order to determine the effect of head gear on the transmission of sound, four standard thicknesses of neoprene (1/8, 3/16, 1/4, and 3/8 inches) will be applied to each head in addition to the condition in which it is tested without covering. Each head will be thoroughly wetted with a detergent before being placed in the water and precautions will be taken to assure that no air bubbles are trapped within the head. The heads will be rotated 180° relative to the projector in 10° steps at each frequency.

Experiment 2d: Transmission of sound energy through the head. In order to determine the influence of the head on the energy loss between the projector and the hydrophone, the head will be rigidly fixed to a calibration station. The projector will be placed 6 feet in front of the midsaggital plane of the head and 6.5 feet from the hydrophone. The calibration station will be arranged so that the head can be rotated relative to the projector and the hydrophone will remain fixed relative to the projector. Measurements will be taken with and without the head in place.

Experiment 2e: Acoustic interaural differences at the cochlea. The experimental arrangement used to determine the IID and ITD as a function of the azimuth of the projector and the type of head covering is the same as the preceding experiment except that in this experiment small (1/2-inch) hydrophones will be placed near the round window of each ear and the difference in arrival time and intensity between the two will be recorded as a function of azimuth of the sound source.

In order to control for the possible changes which immersion in water may make in the tissue over time, the order in which the experiments will be conducted will be counterbalanced--as will the order of frequency and headcovering.

(2) Experiments in Passive Underwater Sound Localization

A number of procedures have been used to determine the subjective azimuth (e.g., pointing, naming an absolute angle, naming a position on a visible scale, drawing the position on a diagram and matching the location with a probe source in the same or another modality). We propose to employ three of these procedures in our experiments.

In the first procedure, subjects will discriminate among fixed sources 45° apart and 8.5 feet distant; a Diver Auditory Localization System (DALS) which is a modification of DICORS (Hollien and Thompson, 1967) will be used in these experiments. DALS is an open framework diving cage constructed of poly-vinyl chloride tubing (PVC tubing is now acoustically invisible underwater); the modifications to DICORS consist of coupling a series of five 8.5-foot arms to the top of the unit. These five arms were located to allow J-9 projectors to be placed at ear level at a distance from the center of the subject's head of 9.5 feet and at angles to the diver/subject of 0°, 45°, 90°, 270°, and 315°. Five projectors will be used to provide the sound sources for the present experiments. In order to calibrate them, an F-36 hydrophone will be fixed to DALS at a position corresponding to the center of the diver's head. The signals from the J-9 projectors will be received by the hydrophone and transmitted by cable to an amplifier (Ithaca model 250) and a divider network on the surface. The signal then will be led to a graphic level recorder (General Radio type 1521-B) coupled mechanically to the beat-frequency oscillator.

Signal voltage and frequency will be monitored by a Voltmeter (Ballantine model 302C), a frequency counter (Hewlett-Packard model 512A), and an oscilloscope. Each of the five J-9 projectors will be calibrated to produce the same SPL reading at the F-36 hydrophone (for all experimental signals) in order to assure that diver/subjects will not be receiving cues on the basis of intensity differences.

The signal source, a Hewlett-Packard wide range oscillator (model 200 CDR) or Grason-Stadler white noise generator (model 455C), will enter a pulse gating unit which will be driven by a pulse timing generator (Scientific Atlanta, series 118) at 1 pulse per second with a duration of 100 to 500 msec. The output of the gating unit will be monitored on a CRO and voltmeter and fed into a booster amplifier which also will be monitored with a CRO and voltmeter. The input to the booster amplifier and its output to the J-9 are controlled by a single switch.

The second procedure (Feinstein, 1971) will require subjects to indicate the position of a sound source which has been moved to any azimuth in 360°. The dependent variable previously was percent correct response to five fixed azimuths while for this procedure the dependent variable will be deviations from the objective azimuth. The "precision" of the localization response will be defined as the correspondence between the objective and subjective azimuth of a single sound source at a constant distance. This procedure was selected because it provided the listener with a response that was: 1) easily made under water; 2) composed of well defined motor behaviors which could be standardized for all subjects; and 3) capable of accurately reflecting the impression of location. The pointing response as defined in these experiments also provides the experimenter with a numerical readout that is accurate to $\pm$ 1° of the indicated position.

For this procedure, the diver sits in a modified DALS; in this instance, however, no arms project from the framework. DALS will be suspended by a special adapter to the center shaft of a unit designed to accurately position sonar domes (and other specialized heavy listening equipment) for calibration purposes. The framework will be secured to prevent twisting and an anchor will be attached to the bottom center of the framework to prevent sway. The sound source (J-9 projector) will be suspended from a boom attached to the outer shaft of the positioning device (PD). The horizontal boom will place the projector face 7.2 feet from subjects' ears and the vertical shaft will drop 14.4 feet from the boom in order to place the projector at ear level.

The projector (outer shaft) and framework (inner shaft) can be rotated independently by means of a servomotor system on the deck. The servomotors are keyed to a Polar Recorder which can be read to the nearest degree. At the beginning of the experiment the inner shaft will be rotated to line up the framework with the long axis of the barge (facing outward) and this shaft will then be locked.

The control switch used by the diver to indicate the azimuth of the source and his confidence on any trial is a plexiglas watertight box filled with transformer oil to prevent fractures due to pressure. In the top center of the box, a flat metal pointer, sharp at one end and rounded at the other, is located beneath a metal shield. The shield is adjusted to allow the subject to grasp the pointer in his right hand with his first finger touching the pointed end; however, this arrangement does not allow the subject to see his hand or the pointer. To the right of the pointer, there are two rows of five buttons and to the left of the pointer are two buttons (which serve as confidence and yes-no indicators). Electrical cable exits the box via Marsh Marine Connectors, dips approximately four meters below the framework and then rises to the control console on the deck. The box will be attached to DALS on hinged arms which will allow the diver to place it in his lap or remove it easily; further, it will be placed so that the pointer is in the median plane of the diver. Calibration to 0° will be accomplished by aligning the pointer with a small upright rod on the box (Subjects feel the point of the rod). The subjective azimuth indicator will be a radio compass indicator (U. S. Army Signal Corps, Indicator I-82-A, Bendix Radio) with an adjustable azimuth and a bank of twelve lights corresponding to the buttons on the diver's control box switch.

Calibration of the course and polar recorder with the median plane of the diver is accomplished visually by dropping a plumb line from the horizontal boom to a marker projecting from DALS. When the plumb and marker coincide, the pen of the polar recorder is set at 0°. Any rotation of the outer shaft can then be read to $\pm 1^{\circ}$ relative to the original starting position.

The third and final procedure to be used provides a measure of auditory localization in terms of minimum audible angles (m.a.a.). In this case, two J-9 projectors are suspended in front of the subjects and the task is to indicate whether the sound comes from the right or the left of the median plane. The angles are preset before each dive by moving two arms on DALS to the appropriate positions. In order to obtain data comparable to earlier underwater m.a.a. studies, the angles used will be 4°, 7°, 18°, 29°, and 40°.

(a) Nonauditory Sound Localization

Experiment 3a: Nonauditory-tactile cues to sound localization (with Dr. JoAnn Kinney, Medical Research Laboratory, U. S. Navy Submarine Base, New London, Connecticut). In order to determine if previous localization performances utilizing the m.a.a. procedure reflected the use of tactile or nonauditory cues, the first experiment carried out utilizing that technique will be replicated: 1) in the presence of an auditory masking tone, 2) with the diver/subjects wearing heavy arctic wet suits, and 3) at stimulus levels below auditory threshold.

Sinusoids of 250, 1000, 6000 Hz and thermal noise will be used as experimental stimuli. The stimulus presentations consisted of five pulses of the experimental frequency set up as 500 msec bursts at 40 dB (110 dB SPL) re: underwater hearing threshold of the poorest hearing subject. The stimulus presentations will be gated ON and OFF with the duty cycle of 500 msec and a 25 msec rise-fall time.

The signals will be presented to diver/listeners five times from each of the five transducers, for a total of 25 presentations of each stimulus. Subjects will respond by means of a five-position underwater switch coupled to an IBM key punch at the surface. Moreover, these responses will be individually verified (by having an assistant check a light panel paralleled to the key punch) before subsequent stimuli are presented. In this manner, errors in recording data will be avoided and subjects will be given ample time to respond to each stimulus presentation. After the subject's response is recorded, a new stimulus will be presented and the procedure will be continued until all 25 presentations of each frequency are completed.

DALS will be lowered by a winch to an ear depth of 40 feet. The diver, wearing open-circuit SCUBA equipment, will descent to the cage, seat himself, lock his arms over a bar provided for subject positioning, and place a lead-weighted belt over his legs to keep him firmly on the seat. During the experiment, subjects will be free to move their head but not their body. A total of ten subjects will be tested (5 male and 5 female). Results should provide information concerning any evidence of tactile underwater localization.

Experiment 3b: Minimum audible angle with no head covering. The first underwater m.a.a. experiments were run in the cold waters of a Nova Scotia bay and therefore subjects had to wear wet suit hoods. The hoods were of 3/16-inch neoprene with holes over the ears. This experiment will be replicated in the warmth and quiet of a Florida spring with and without hoods.

(b) Are Binaural Difference Cues Determining Underwater Sound Localization?

Experiment 3c: The effect of reorganization of binaural cues in air on subsequent localization in water. This experiment will determine whether the same cue hierarchy is operating to determine sound localization in both air and water. Other investigators have found that when subjects are made to wear an attenuating ear muff over a reasonably short period of time they first localize sound toward the uncovered ear and gradually they normalize their responses so that the sound source is eventually localized in the correct position. Upon removal of the muff, the subject generally exhibits an after-effect which causes the sound to be localized toward the previously covered ear. The subjects in this experiment will be required to wear an attenuating muff for a period of 72 hours and they will be tested at 24 hour intervals (10 trials per position with no feedback). After the last test, one-half of the divers will remove their muff and enter the water to complete another series of localization tests whereas the other one-half will be tested only in air. The localization precision procedure will be used in this experiment.

(c) Phantom Source Experiments in which Time, Intensity, and Phase of Multiple Sources are Manipulated

The experiments proposed in this section will utilize a combination of the first two of the three procedures described above. DALS will be used to present sound sources in fixed locations but the subjects will respond by indicating the apparent location of the phantom source using the pointing system (with Dr. D. C. Teas, C.S.L.). The applicants are aware that phantom sources are typically studied using earphones to avoid confounding interaural difference cues. The following experiments are proposed with that limitation in mind and appropriate controls are planned.

Experiment 3d: Phase determined phantom source. In this experiment, manipulation of the apparent direction of the sound source will be attempted by means of the Teas-Jeffress technique in which multiple sound sources are used with signals in various phase relationships. This project will assist in the identification of the auditory component in underwater sound localization as the nature of the data will, in a general way, be comparable to those obtained by Teas in air. For example, if they agree, the data will argue an auditory component in underwater localization. Ten subjects will be studied.

Experiment 3e: Location of the phantom source determined by intensity imbalance. Equal onset time with 20 attenuations from 1/10 to 2 dB programmed to occur randomly on either the right or left speaker, will be used to shift the phantom source back and forth across the midline. The subjects will respond by pointing at the apparent location of the source; ten individuals will be employed as subjects.

Experiment 3f: Location of the phantom source determined by lead time (Precedence Effect). Equal intensity with lead time on one side or the other ranging from 0 to 200 msec (10 settings 20 msec apart), will be used to shift the phantom source back and forth across the midline. The subjects will respond by pointing at the apparent location; ten divers will be used as subjects.

Experiment 3g: The time vs. intensity trading ratio in sound localization underwater. The final experiment of this series will delineate the relationship between intensity and time difference cues on underwater localization. In the first 20 trials, the phantom source will be centered by balancing the intensities of the two speakers. The intensities will then be unbalanced and the onset time will be manipulated to return the phantom to the center position. The method of limits will be used to determine the point at which the time cue balances the intensity cue. The experiment will be replicated with the intensity cue used to center the source when the time cue is unbalanced.

d. Significance of the Research

Currently, tremendous strides are being made with respect to life support systems for divers. This situation, coupled with the need for more working divers of all types, and advances in recreational diving, is resulting in the explosive expansion of the numbers of individuals engaged in this activity (already there are over .5 million divers in the U. S. alone). Thus, we are at a point in time where the development of good communications and navigation techniques are vital--as is the development of mechanical and electronic aids to diver communication and navigation. In order that such development can take place, precise and reliable information about underwater auditory function is necessary. Hence, included among the research thrusts critically needed at present is a major program in underwater sound localization research. Our proposed program is focused on this very issue.

Some practical examples concerning the need for information of the type our project would provide are as follows: A very important aspect of underwater communication is the transmission of information about the relative location of divers, e.g., knowledge of the location of a buddy diver may be the difference between survival and drowning. The designation of effective signals and design of appropriate signalling devices of a type that will allow such specification depends on the data we propose to obtain. Further, the kinds of transducers and their configurations presently used in diver communication have not been developed from empirical data, yet highly efficient units must become available if adequate communication is to be accomplished. The experiments which we have proposed will provide information about the modes and mechanisms of auditory function underwater and the capacity of divers to respond to the process, time and intensity information. In short, what the diver can hear--and how well he hears it--are important both to his ability to communicate and to navigate in an essentially hostile environment.

Underwater navigation is as imp tant to the diver as are good communications. If we take our cue fro e marine mammals, acoustic navigation may be an effective remedy for the absence of visual cues in the sea. In this regard, we have considered such an alternative in our earlier research on sound localization and have determined that reasonable degrees of acuity and precision are obtainable underwater. We also have inferred that this phenomenon of localization may be translatable into a capacity for navigation to a distant sound source.

In addition to the several justifications provided above and those inherent in proposed research, we should like to point out certain unique features with respect to the applicants and their research environment. We are both trained working divers; we both have developed (independent) research <u>programs</u> in underwater sound localization (to our knowledge, we are the only workers who have done so) and we are pooling our efforts. Further, we enjoy rather unique support as follows: 1) Both our specified (Dr. G. C. Tolhurst and Dr. J. Zwislocki) and unspecified (Dr. D. C. Teas and Dr. G. Bond) consultants provide a spectrum of general expertise and specific knowledge of our problems that is virtually unrivaled; 2) We have readily available an excellent population of highly trained and/or scientific divers from CSL and the U. S. Navy; 3) We have available (via the CSL holdings and from the cooperating Navy laboratories) virtually all of the specialized equipment necessary for the successful conduct of our research program; and 4) We have available to the project the support of a number of outstanding Navy laboratories including NRL's Underwater Sound Reference Division, Orlando, Florida, and Naval Ship Research and Development Laboratory, Panama City, Florida.

References

Albers, V. M., Underwater Acoustics Handbook, University Park, Penn.: The Penn. State Univ. Press, Vol. II, 1965.

Andersen, S. and Christensen, H. T., Underwater sound localization in man, J. Aud. Res., 9, 358-364, 1969.

Batteau, D. W., The role of the pinna in human localization, Proc. Roy. Soc. (Biol.), 168, No. 1011, Series B., 158-180, 1967.

Batteau, D. W., Listening with the naked ear, In Freeman, S. J. (Ed.), The Neurophysiology of Spatially Oriented Behavior, Homewood, Ill.: The Dorsey Press, 1968.

Bauer, B. B. and Torick, E. L., Experimental studies in underwater directional communication, J. Acoust. Soc. Amer., 39, 25-34, 1966.

Brandt, J. F. and Hollien, H., Underwater hearing thresholds in man, J. Acoust. Soc. Amer., 42, 966-971, 1967.

Brandt, J. F. and Hollien, H., Underwater hearing thresholds in man as a function of water depth, J. Acoust. Soc. Amer., 46, 893-894, 1969.

Corso, J. F., Bone conduction thresholds for sonic and ultrasonic frequencies, J. Acoust. Soc. Amer., 35, 1738-1743, 1963.

Corso, J. F. and Levine, M., The pitch of ultrasonic frequencies heard by bone conduction, Proc. Penna. Acad. Sci., 37, 22-26, 1963.

Corso, J. F. and Levine, M., Pitch discrimination at high frequencies by air and bone conduction, Amer. J. Psychol., 78, 557-566, 1965.

Corso, J. F. and Levine, M., Sonic and ultrasonic equal-loudness contours, J. Exp. Psychol., 70, 412-416, 1965.

Dudok van Heel, W. H., Audio-direction finding in the porpoise (Phocaena phocaena), Nature, 183, 1063, 1959.

Fedderson, W. E., Sandal, T. T., Teas, D. C. and Jeffress, L. A., Localization of high frequency tones, J. Acoust. Soc. Amer., 19, 988-991, 1957.

Feinstein, S. H., Human hearing underwater: Are things as bad as they seem? J. Acoust. Soc. Amer., 40, 1561-1562, 1966.

Feinstein, S. H., Underwater sound navigation, Unpublished paper given at Man's Performance in the Sea, Seminar, Communication Sciences Laboratory, University of Florida, 1969.

Feinstein, S. H., The acuity and precision of underwater sound localization, Doctoral Dissertation, Dalhousie University, Halifax, Nova Scotia, 1971.

Firestone, F., Phase differences and amplitude ratio at the ears due to a source of pure tone, J. Acoust. Soc. Amer., 2, 260-270, 1930.

Gentry, R. L., Underwater auditory localization in the California sea lion (Zalophus californianus), J. Aud. Res., 7, 187-193, 1967.

Green, D. M. and Henning, G. B., Audition, Am. Rev. Psychol., 20, 105-128, 1969.

Haeff, A. V. and Knox, C., Perception of ultrasound, Science, 139, 590-592, 1962.

Hamilton, P. M., Underwater hearing thresholds, J. Acoust. Soc. Amer., 29, 792-794, 1957.

Hartley, R. V. and Fry, T. C., The binaural localization of pure tone, Phys. Rev., 18, 431-442, 1921.

Hollien, H., Underwater sound localization: Preliminary information, J. Acoust. Soc. Amer., 46, 124-125, 1969.

Hollien, H., Lauer, J. L. and Paul, P., Additional data on underwater sound localization, J. Acoust. Soc. Amer., 27, 127-128, 1970.

Hollien, H. and Brandt, J. F., The effect of air bubbles in the external auditory meatus on underwater hearing thresholds, J. Acoust. Soc. Amer., 46, 384-387, 1969.

Hollien, H. and Thompson, C. L., A Diver Communication Research System (DICORS), Progress Report CSL/ONR No. 2, ONR Grant Nonr 580(20), 1-5, AD-648, 935, 1967.

Hollien, H., Stouffer, J. L., and Lauer, J. L., Additional data on underwater sound localization in man, Unpublished manuscript, 1971.

Ide, J. M., Signaling and homing by underwater sound for small craft and commando swimmers, NRL Sound Report No. 19, June, 1944.

Jeffress, L. A. and Taylor, R. W., Lateralization vs. localization. J. Acoust. Soc. Amer., 33, 482-483, 1961

Kellogg, W. N., Porpoises and Sonar, Chicago: Univ. of Chicago Press, 177 pp., 1961.

Kellogg, W. N., Sonar system of the blind, Science, 137, 399-404, 1962.

Leggiere, T., McAniff, J., Schenck. H., and van Ryzin, J., Sound localization and homing in divers, Mar. Tech. Soc. Jour., 4, 27-34, 1969.

Luria, S. M. and Kinney, J. A. S., Underwater vision, Science, 167, 1454-1461, 1970.

Miles, S., Underwater medicine, London: Staples Press, 2 editor, 332 pp., 1966.

Mills, A. W., On the minimum audible angle, J. Acoust. Soc. Amer., 30, 237-246, 1958.

Mohl, B., Auditory sensitivity of the common seal in air and water, J. Aud. Res., 8, 27-28, 1968.

Nordlund, B., Physical factors in angular localization, Acta oto-laryng., 54, 75-93, 1961.

Nordlund, B., Angular localization, Acta oto-laryng., 55, 405-424, 1962.

Norris, K. S., Some problems of echo location in cetaceans, in Tavolga, W. N. (Ed.), Marine Bio-acoustics, New York: Pergammon, Vol. 1, 1964.

Rayleigh, Lord, (J. W. Strutt), On our perception of sound direction, Phil. Mag., 13, 214-232, 1907.

Reysenbach de Haan, F. W., Hearing in whales, Acta oto-laryng., Suppl., 134, 1-114, 1957.

Sagalovich, B. M. and Melkumova, G. G., New data on the spectrum of sound and ultrasound frequencies producing sound effects in man, Res. Insti. of Ear, Nose and Throat Diseases, Moscos, 1967.

Sandel, T. T., Teas, D. C., Fedderson, W. E., and Jeffress, L. A., Localization of sound from single and paired sources, J. Acoust. Soc. Amer., 27, 842-852, 1955.

Sedee, G. A., On spatial binaural hearing, Thesis, Utrecht, 195[illegible].

Shaxby, J. H. and Gage, F. H., Studies in the localization of sounds in the median plane, Sp. Rev. Serv. Med. Res. Coun. Loud., 166, 1-32, 1932.

Sivian, L. J., On hearing in water vs. hearing in air, J. Acoust. Soc. Amer., 19, 461-463, 1947.

Stevens, S. S. and Newman, E. B., The localization of pure tones, Proc. Nat. Acad. Sci., Wash., 20, 593-596, 1934.

Stevens, S. S. and Newman, E. B., The localization of actual sources of sound, Amer. Jour. Psychol., 48, 297-306, 1936.

Tobias, J. V. and Zerlin, S., Lateralization threshold as a function of stimulus duration, J. Acoust. Soc. Amer., 31, 1591-1594, 1959.

Woodworth, R. S., Experimental Psychology, New York: Holt, 1938.

Zerlin, S., Personal communication, 1969.

Howard B. Rothman

I. Personal Research

1. Introduction: Research Program in Acoustic Phonetics

The study of acoustic phonetics provides a precise and reliable means of defining an encoded event and relating it to a decoded event. This relationship provides a link between speech production and perception. Recent studies have shown that an acoustic signal provides more information than is needed for perception; many of these cues are highly redundant. The techniques of acoustic phonetics has enabled investigators to specify those portions of the acoustic signal which are primary cues for speech perception. The significance of this type of research has broad application to many areas, e.g., the development of speech synthesizers, high volume tele-communications and automatic speech translators and reading machines.

The acoustic signal reflects a highly complex and synergistic relationship between the structures of speech varying over time. Relating changes in the acoustic signal to structural changes and movements has broad application to the area of speech pathology. Speakers with normal speech and hearing mechanisms learn and develop specific physiological and anatomical relationships which are not universals of all languages but which are necessary within a particular language community. The absence or alteration of these relationships causes speech production that is different from the norm or standard of the community, hence causing a perceptual difficulty. Better therapeutic techniques based upon quantifiable and specific measurements can be developed by comparing and contrasting normal and aberrant acoustic signals especially since direct physiological study is difficult.

Greater understanding of the complex and interactive processes comprising speech communication can be gained by combining the advantages of acoustic techniques with those of physiologic methods. The Communication Sciences Laboratory has the expertise and equipment to do this. Because of this unique situation I have been able to design a research program that uses acoustic and physiologic techniques for investigating normal, pathologic and exotic phenomena of speech.

2. An Acoustic and Electromyographic Analysis of Consonant-Vowel Transitions in the Speech of Deaf Adults

a. Introduction

The process of speech communication is one of the most intricately integrated functions of humans. Yet, communication via speech is universal in humans and is most often acquired without formal instruction by people with normal hearing. On the other hand, the congenitally deaf and those who acquire severe hearing losses at an early age can acquire speech only through formal instruction that stresses conscious control of a motor skill. Deaf children must rely on visual, tactile, and kinesthetic cues to develop coordination and control of groups of widely distributed muscles and functions. Normal children develop these skills by responding to the sounds they hear themselves making and by response to sounds in their environment.

There is general agreement among those who work with deaf people that deaf speech is different, strange and difficult to understand. Many of the characteristics which cause this strangeness have been identified. Among them are articulation errors which include both consonants and vowels, and rhythmic errors which involve durational distortions within and between phonemes. In spite of the agreement that there is an entity identifiable as deaf speech there remains the need to study the acoustic and physiological components of its characteristics so that clinicians and teachers of the deaf can deal with them effectively.

b. Acoustic Studies of Normal Speech

This section does not intend to present an exhaustive review of all literature pertaining to deaf or normal speech. Rather, it will present the results of some of the major and pertinent studies dealing with the speech of the deaf as well as results of acoustical studies dealing with speech production in normal hearing individuals, specifically related to this study.

A few investigations of deaf speech imply that intelligibility is often affected by a speaker's attempts to combine relatively discrete and invariant articulatory responses into a continuously varying acoustic event. However, research conducted by workers at Haskins Laboratory (Delattre, Liberman and Cooper, 1955; Delattre, Liberman and Cooper, 1964; Harris, Huntington and Sholes, 1968; MacNeilage and Sholes, 1964) by Lindblom (1963) and Ohman (1966a) have shown that the articulatory gesture often associated with a given phoneme varies with the phonological context, and that a single acoustic cue carries information in parallel about preceding and successive phone segments. Much of this parallel information is carried by the acoustic transitions between phoneme segments which represent articulatory movement from--or to--the place of consonant production to--or from--the position of the adjacent vowel (Delattre, Liberman and Cooper, 1955). The parallel delivery of information results in a complex relation between acoustic cue and perceived phoneme; that is, the cue for a particular phoneme will be different for each context (Liberman, Cooper, Shankweiler and Studdert-Kennedy, 1968).

The transition between a consonant-vowel (CV) pair is an intrinsic part of that specific CV pair and cannot be recombined with another CV pair or with the same CV pair produced in a different context. Cyril Harris (1953) tried to recombine phonemes through tape splicing and was unsuccessful because of the overlapping in time and the mutual influence (the effect of coarticulation) of the production of the particular phonemes of a specific CV pair in the specific context. The reason for the coarticulation of phonemes is due to the physiological constraints imposed by instructions to the articulators occurring in close temporal succession (Lindblom, 1963) and not because of differences in speaker intention.

Lindblom (1963) found that stress, rate of utterance and contextual influence caused vowel reduction, i.e., vowel formant frequencies are characterized acoustically by undershoot or a failure to reach their ideal steady-state frequencies. If signals to the articulators are far apart in time target values may be reached. However, in connected discourse, where the articulatory system may be responding to several signals simultaneously and there is less time for completion of a movement towards a target value one normally expects vowel reduction to occur. This vowel reduction results in the centralization or neutralization of vowels.

Lindblom's contention that vowel reduction is a normal part of connected discourse supports Stevens and House (1963) who found that consonantal context caused a shift downward for front vowels which have a high second formant (F_2), and a shift upward for back vowels which have a low F_2. The physiological effect of this acoustical change is the shifting of a vowel's tongue position towards the point of constriction for an adjacent consonant. So when a production of a particular vowel requires a constriction in the vocal tract that is remote from the place of articulation for an adjacent consonant, e.g., from a low back vowel to a post-dental consonant, undershoot in the motion of the articulator during the vocalic position of the syllable could result in a vocal tract configuration that is less constricted than that of the ideal target for the vowel (Stevens and House, 1963, p. 125). This displacement results in an acoustical shifting of formants towards the schwa.

Stevens and House (1963) also found that F_2 values for vowels in a fricative environment were relatively lower for front vowels and relatively higher for back vowels than the corresponding values for a stop consonantal environment. This finding indicates that vowels in a fricative environment tend to shift further from their ideal target configurations than do the same vowels in stop environments even though vowel durations for fricative environments are longer. There is no contradiction between the latter statement and Lindblom's (1963) contention that faster rates of utterance resulting in signals to the articulators occurring in close temporal succession cause greater shifts away from a vowel's target values. One has only to consider that articulatory structures can execute displacements to and from the complete closures necessary for stop consonants at a much greater speed than they can with fricatives which require deceleration to, and acceleration from constrictions of precisely controlled size and shape (Stevens and House, 1963).

As indicated by the above studies, speech depends upon the synergistic action of the articulators. Additional evidence that a series of phonemes put together in a VCV utterance cannot be considered as a linear series of discrete adjacent sounds or gestures is found in Ohman's study (1966a) of the effect of three Swedish rounded vowels across the intervocalic closure for /b, d, g/. By changing one vowel at a time, keeping everything else constant, Ohman was able to study the effects of the changing element across the intervocalic closure. As shown in Figure 1, changing the final vowel has a great effect on the formants and transitions of the initial consonant-vowel across the intervocalic stop. In the first case, F_2 preceding the stop rises when the final vowel is /y/ and, in the second case, falls when the final vowel is /ɑ/. The method of observing the influence of a changing element on a constant frame has been adapted for use in this study because it provides a set of minimally different pairs. It is often difficult to identify formant transitions or to decide where a transition begins by examining a single spectrogram. By looking at sets of spectrograms of minimally different pairs, transitions can be defined with respect to a set of entities that are otherwise identical and differences in formant transitions can be pointed out if they exist (Halle, Hughes and Radley, 1957).

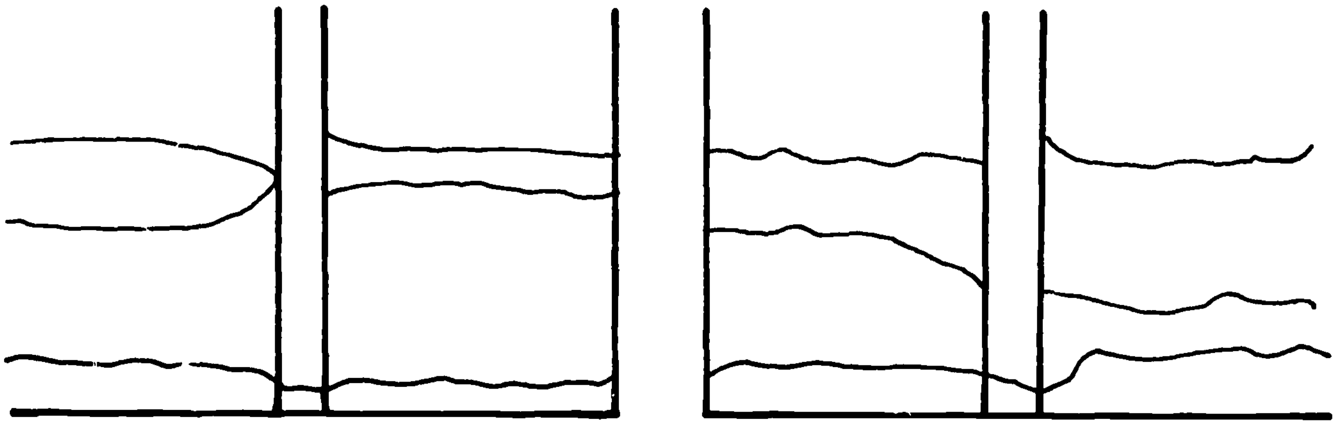


Figure 1. Spectrograms of two utterances spoken by a Swedish male. The F_2 transitions preceding the stop are different due to the influence of the different final vowels (Ohman, 1966a).

The investigations described above have amply demonstrated the synergistic action of the articulators. The overlapping of adjacent gestures cause a single acoustic cue to carry parallel information about successive phonemic segments. Called coarticulation, the effect of parallel delivery of information enables human perception to overcome the limitations of the temporal resolving power of the ear and to perceive speech as fast as it does. Because of coarticulation it is often impossible to divide the acoustic signal in such a way as to recover a segment that will stand alone as a particular phoneme. Also, due to the overlapping in time of an articulatory sequence one can see the influence of a vowel across a consonant gesture.

c. Acoustic Studies of Deaf Speech

Studies of deaf speech carried out by Hudgins and Numbers (1942), John and Howarth (1965), and Hood (1966) have stressed the fact that the deaf treat phonemes, syllables and words as isolated events rather than as integral parts of a changing larger whole. Because the deaf tend to treat segments of an utterance as discrete events there tends to be an absence of the coarticulation effect (Ohman, 1966a).

Hudgins and Numbers (1942) found the primary reason for low intelligibility in the speech of the deaf to be articulation and rhythmic errors. By grouping the error categories involved into two groups on the basis of motor processes, two main errors emerge. They are the inaccuracy or failure of articulation and the lack of coordination between the several component muscle groups comprising the complex speech mechanism. These authors further report misarticulation of compound consonants (blends), diphthongs and vowel neutralization to be categories most highly associated with poor intelligibility and to contain the greatest number of errors. All of these categories require movement of the articulators and should show the presence or absence of coarticulation.

Hudgins and Numbers (1942), John and Howarth (1965), and Hood (1966) all found that abnormalities of speech rhythm are highly correlated with poor intelligibility. Hood (1966) also found that listeners were able to differentiate normal speakers from deaf speakers on the basis of speech rhythm proficiency only. John and Howarth (1965) demonstrated that improving the time aspects of deaf speech, i.e., improving temporal factors and continuity, resulted in a significant improvement in intelligibility. Though John and Howarth do not directly say why improving the continuity between syllables and words and discouraging pauses between words improved intelligibility it is obvious that speech segments which were tested as isolated events became integral parts of a changing larger whole whose components mutually modify each other. By eliminating intrusive sounds and silences between words and by speaking faster, the articulation of these deaf speakers most probably responded to the physical-mechanical constraints of the articulatory system which resulted in coarticulation.

Another factor warrants discussion. Both Hood (1966) and Calvert (1961) found durational differences between deaf and normal groups. In Hood's study the deaf group with good speech rhythm spoke twice as slowly as the normals and the deaf group with poor speech rhythm spoke $3\frac{1}{2}$ times as slowly as the normals. Calvert found that the deaf had longer mean closure durations for voiced consonants than for voiceless consonants which is a reversal of the trend seen in normals as reported by Lisker (1957). The only aspect of phoneme duration in Calvert's deaf group that followed the pattern seen in normals was that of the release period for stops. The release period for voiceless plosives was greater than for voiced plosives as would be expected due to the greater build-up of transglottal pressure (Halle, Hughes and Radley, 1957).

In addition, Calvert (1961) also conducted a survey among teachers of the deaf in the San Francisco Bay Area which indicated that deaf speakers have a characteristic "voice quality" that can identify the speaker as being deaf. Although there was a wide divergence of opinion as to the specifications of this quality, Calvert found that some articulatory movement was necessary before his listeners consistently identified the speakers as being deaf or normal. The necessity of some articulatory movement before consistent identification of a speaker as being deaf was confirmed by Hood (1966), who found that it is primarily in speech over time that a deaf speaker's true phonatory characteristics emerge. Calvert's results indicating that the quality identified as deaf speech may be durational distortions involving relational values between phonemes as well as absolute duration value of phonemes is consistent with the results of the investigations discussed above.

The investigations discussed above agree that the speech of the deaf does not reflect the reciprocal durational effects of adjacent phonemes. This suggests that their phonemes are poorly joined and are treated as discrete units and not as part of connected, mutually influenced series of connected events. By teaching the production of speech as a series of static events, e.g., stressing the articulation of isolated phonemes, one introduces a series of distortions into the speech process. These distortions include those of rhythm, stress and duration of single phonemes and of durational relationships between phonemes. The evidence presented by the above studies indicates that once an adequate level of articulation for isolated phonemes has been established, emphasis should be placed on establishing certain minimum levels of speech rhythm. Thus, by maintaining correct speech rhythm the speech of the deaf should reflect a response to the physical-mechanical constraints of the articulatory system which would then result in an increase of coarticulation.

d. Electromyographic Studies of Normal and Deaf Speech

Electromyography (EMG) is a technique that is well suited to the study of speech (Cooper, 1965; Gay, 1968). Unlike spectrograms which supply indirect information about speech gestures, EMG enables one to look directly at muscle action potentials while the muscles involved are performing the skilled movements necessary for speech. Muscle action potentials are recorded by electrodes, placed on or near a muscle, which pick up the electrical potentials given off by muscular contraction. In the study of speech gestures where one is concerned with overall muscle activity rather than with the activity of a single motor unit, investigators have used surface electrodes. Surface electrodes are easily placed, are held on by paste or suction, and cause minimal discomfort over substantial periods of time. Placing the surface electrodes at different positions can give much information on the component parts of a speech gesture as well as information on the timing and force of the components. Descriptions of the various types of systems devised can be found in the literature (Blinn, 1955; Harris, Rosov, Cooper and Lysaught, 1964; Moore, 1966).

MacNeilage and Sholes (1964) studied muscle action in the tongue of normal speakers by placing surface electrodes, three at a time, in thirteen positions on the tongue. They used only one subject who repeated monosyllables of the shape /pVp/. Cinefluorograms were taken of the vowel production so tongue position could be studied. MacNeilage and Sholes showed that surface electrodes were able to pick up muscle action potentials; to specify, based on their one subject, which muscles produced observed tongue shapes, and that a direct relationship exists between voltage levels and the amount of tongue movement.

MacNeilage and DeClerk (1969) used surface EMG to study coarticulation during CVC monosyllables. Using one subject they recorded 36 CVC syllables. Surface electrodes were placed on the upper articulatory musculature and on the tongue. The investigation of MacNeilage and DeClerk showed that EMG can be used to study coarticulation. Left-to-right effects were seen in all cases and right-to-left effects were observed in most cases thus indicating that the CVC syllable is not composed of a series of context-independent phonetic elements. MacNeilage and DeClerk maintain that the CV pair is more cohesive than the VC pair. Contextual modifications, they say, are the result of the changes caused by the articulatory system's handling of discrete voluntary commands which originate at a neurological level above that which serves a specialized motor function.

Huntington, et al. (1968) used surface EMG to study the topology, i.e., the articulatory configurations, in two deaf speakers and two normal speakers. Surface electrodes were placed on the lips and tongue while the subjects read a series of utterances of the form /hɑCVk/. Their results show that the deaf speakers behaved differently than did their normal speakers, and that variability within groups (between speakers) was higher for the deaf than for the normal speakers. In addition, their results show that deaf speakers have definite patterns of articulation which are not comparable to those of normals.

As indicated by some of the above studies, surface EMG is a useful way of looking at topological features of speech. None of the above studies have utilized this technique for examining transitions. The movement of transitions as seen in acoustic records is a product of changes in resonance brought about by variations in cavity size effected by the moving articulators. Surface EMG provides an indication of the force or contraction of many motor units of a muscle movement. Therefore, the record of muscle potential is not comparable in any simple manner to changes in cavity resonance. Acoustic transitions reflect variations in cavity size due to articulatory movement. However, surface EMG may provide information about the nature or timing of the command to a muscle or articulator for a given phoneme in different contexts.

e. Statement of the Problem and Purposes

There is no question that deaf people have great difficulty communicating through speech, i.e., oral language. The deaf would have trouble talking, and when they do talk they sound different from normals and are often identified as being deaf speakers. Few studies have looked at the various elements of deaf speech to identify those elements that are most troublesome for the deaf, and to specify which are most highly correlated with poor intelligibility.

Methods of teaching and training often appear to compound the inherent difficulties that deaf speakers have with oral language. The teaching of speech to the deaf should concern itself with developing an intelligible medium of communication. Two generalized approaches are used for teaching speech to the deaf. These are: 1) the analytic method and 2) the synthetic method.

The analytic or elements method stresses the teaching of individual speech sounds. The deaf child is taught to imitate individual and isolated phonemes which are then combined into words. The words are then combined to form phrases. Each phoneme is given a relatively fixed articulatory position, and relative durational, qualitative and intensity values before they are combined into syllables and words.

The synthetic method stresses the imitation of whole words before the specific elements comprising a word have been learned. Phonemes are worked on and corrected in the context of words. Movement of the articulators is stressed over articulatory position. Easy and natural sounding speech is the goal of the synthetic method. Unfortunately, in the speech of most deaf speakers, the synthetic method breaks down and analytic components are more evident than not.

Many perceptual and acoustic studies (discussed above) have shown that individual phonemes, regardless of how correctly they are articulated, account for only a part of the intelligibility of speech, and that phonemes in context bear little relationship to phonemes in isolation. Unless there exists a proper relationship between phonemes in sequence, speech will be no more intelligible than are the results of synthesizing speech by stringing together isolated phonetic elements. Cyril Harris (1953) argues that individual phonemes of a word cannot be separated and then respliced back into different combinations without greatly reducing intelligibility. Harris did not take into account the fact that the anticipation of an oncoming vowel during the production of a consonant is an integral feature of normal speech. That it is far less a feature of deaf speech is probably due to training techniques. Because speech is a dynamic process involving the close relationship and synergic action of different muscles and muscle groups one must teach and stress speech as a dynamic process (Hudgins and Numbers, 1942).

As noted previously, the synergistic action of the articulators produces coarticulation. Due to coarticulation, and the overlapping in time of an articulatory sequence, one sees the influence of a vowel across consonant gestures. This influence of a vowel across consonant gestures may be considered a part of normal speech. It is the intention of this investigator to examine the transitions of deaf speech, the presence or absence of coarticulation effects and the neutralization of vowels in order to determine whether the influence of overlapping articulatory events will be seen in the speech of the deaf.

Much of deaf speech training consists of phoneme, syllable and word articulation practice, many deaf speakers have had a great deal of practice with small elements and with the reading of word lists. The influence of the physical-mechanical constraints imposed upon the articulatory mechanism over time by a changing phonetic environment is a part of all real speech situations. These constraints may not be as prominent or may be modified in an artificial situation such as a word list. Since it is primarily in speech over time that a deaf speaker's true phonatory characteristics emerge it is the intention of this investigator to study certain aspects of the speech of deaf speakers by approximating a real speech situation as closely as possible while still maintaining control over various parameters.

f. Subject Selection

As stated, the purpose of this investigation will be to study the differences between the articulatory behavior of adult normal speakers and adult deaf speakers in a controlled phonetic context. Accordingly, at least eight to ten adults will be ultimately chosen for each group--for a total of 16 to 20 subjects.

The criteria to be used for choosing the deaf population will include: 1) presence of long-term profound bilateral hearing loss, 2) similar speech training backgrounds, and 3) the ability to articulate the stimulus items adequately. The criteria to be used for choosing the normal hearing population will be: 1) no speech problem; 2) no hearing loss, and 3) that they all be representative of a single dialect area.

A screening procedure utilizing teachers and clinicians of the deaf will be used to categorize the deaf subjects as to type of hearing loss and ability to articulate the stimulus items adequately.

Stimulus items will be chosen in order to emphasize differences in formant transitions and will consist of minimally different pairs of items embedded within the sentence "Take a ______ aside." Four consonants (/t, k, l, s/) in initial position and three vowels (/i, ɑ, u/) will be used to form monosyllabic words with /t/ always being the final consonant of the word. The voiced stops, /d, g/ may also be used to investigate the additional problem deaf speakers have in coordinating voicing with articulation. Taking the words of a list and embedding them in a sentence frame provides a linguistically real situation in which to judge various parameters, and the influence of the changing medial key word on the constant sentence frame can be investigated.

It should be noted that all of the phonemes represent either a contrasting articulation or an articulatory extreme. In the case of the vowels /i, a, u/, the representation is of the articulatory extremes of the traditional physiological vowel chart. Since there is a great deal of acceptable allophonic variation in vowel articulation and because the effect of diminished stress and context effect a neutralization or undershoot of a vowel towards the schwa (Lindblom, 1963; Stevens and House, 1963), it was felt that an articulatory extreme would especially help the deaf speakers achieve good representations of each vowel in the various contexts. In addition to the articulatory extremes of height (cephalad-caudal) and depth (posterior-anterior), the juxtaposition of /u/ with /i/ represents the contrast of a lip-rounded vowel with a lip-spread vowel.

The consonants provide alveolar and velar stop (/t, k/) and sustained (/l, s/) contrasts. The production of /k/ is especially variant when paired with a front or back vowel due to the different anterior-posterior points of closure (Halle, Hughes and Radley, 1957). Unlike the stop consonants where articulatory structures can execute movements to and from a point during a closure, the production of /s/ requires a sustained precise control of the articulators over time (Stevens and House, 1963). The juxtaposition of /t/ and /k/ with /s/ may show differences in the way deaf and normal speakers approach a vowel in a stop or fricative environment where, in the one case, a greater degree of parallel articulation can occur in a context of shorter duration. The consonant /l/ provides a large degree of continuous articulation, and therefore, of sustained formant transition during which coarticulation effects should be seen.

The /t/ was chosen as the final element for all of the twelve key words because it provides a good kinesthetic articulatory reference point and has a more stable acoustic locus when juxtaposed with different vowels (Delattre, Liberman and Cooper, 1955). In addition, the stop articulation of the /t/ provides a fairly precise and rigid articulatory reference point for the transition to the following vowel.

Each consonant will be combined with each vowel to give a set of twelve key items. These items are: toot, teet, tot; coot, keet, cot; lute, leet, lot; and suit, seat, sot, and they form sentences such as "Take a TOOT aside." It is the effect of the changing key word on the surrounding sentence that will be investigated.

The key words in their sentence frame will be presented to the subjects ten times each in random order resulting in 120 presentations. Multiple utterances of each stimulus item are felt to be necessary in determining what spectrographic segment constitutes a transition and where a transition begins or ends (Halle, Hughes, Radley, 1957; Ohman, 1967). In addition, for the interpretation of EMG data, multiple utterances are necessary to overcome both the effects of movement artifacts and the natural variability between tokens of an utterance.

Simultaneous recordings will be made of two types of data: spectrographic and electromyographic. A multi-channel FM tape recorder will be used to facilitate the comparison between EMG traces and acoustic events in real time. Electrodes will be placed in two positions on the tongue as well as in various positions on the lips. These positions should give a picture of different aspects of the articulatory process, though not a complete one. A simple, inexpensive, yet effective multiple suction electrode system has been designed by this investigator and will be built at the Communication Sciences Laboratory (Figure 2).

The type of study outlined above has particular relevance to the multidisciplinary milieu of the Communication Sciences Laboratory. Specifically, the multiple suction electrode system can be used to augment studies of lingual and intraoral air pressures currently under investigation by other members of the Laboratory. A simultaneous investigation of muscle action potentials with air and tongue pressure variances will add to the understanding of the basic and complex process of speech articulation.

References

Angelocci, A.A., Kopp, G. A., and Holbrook, A., The vowel formants of deaf and normal-hearing eleven- to fourteen-year-old boys, J. Speech Hearing Dis., 29, 156-170, 1964.

Blinn, Kenneth A., A suction EMG electrode assembly, Electroenceph. Clin. Neurophysiol., 7, 141-142, 1955.

Calvert, D.R., Some acoustic characteristics of the speech of profoundly deaf individuals, Ph.D. Dissertation, Stanford University, 1961.

Cooper, Franklin S., Research techniques and instrumentation: EMG, ASHA Rep., 1, 153-168, 1965.

Cooper, F., Delattre, P., Liberman, A., Borst, J. and Gerstman, L., Some experiments of the perception of synthetic speech sounds, J. Acoustic Soc. Amer., 24, 597-606, 1952.

Delattre, P.C., Liberman, A.M. and Cooper, F.S., Acoustic loci and transitional cues for consonants, J. Acoust. Soc. Amer., 27, 769-773, 1955.

Delattre, P.C., Liberman, A.M. and Cooper, F.S., Formant transitions and loci as acoustic correlates of place of articulation in American fricatives, Studia Linguistica, 104-121, 1964.

Denes, P., Effect of duration on the perception of voicing, J. Acoust. Soc. Amer., 27, 761-764, 1955.

Gay, T., Current status of electromyography in speech research, SR-15/16, 125-136, 1968.

Grossman, W.I. and Weiner, Herbert, Some factors affecting the reliability of surface electromyography, Psychoso.Med.,28, 78-82, 1966.

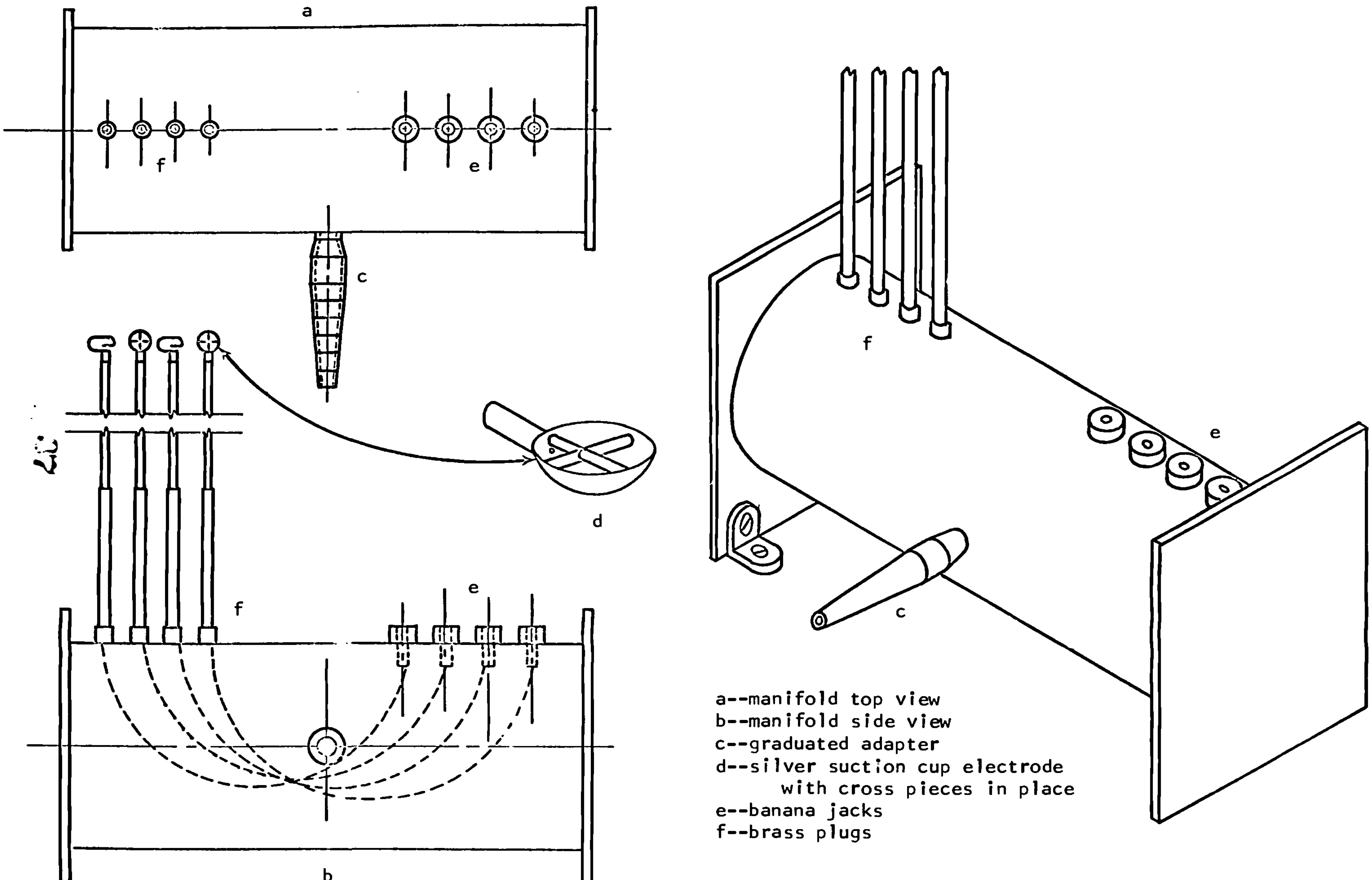

a
b
c
d
e
f
a--manifold top view
b--manifold side view
c--graduated adapter
d--silver suction cup electrode
with cross pieces in place
e--banana jacks
f--brass plugs

Figure 2.

Halle, M., Hughes, G.W. and Radley, J.P.A., Acoustic properties of stop consonants, J. Acoust. Soc. Amer., 29, 107-116, 1957.

Harris, C., A study of the building blocks of speech, J. Acoustic Soc. Amer., 25, 962-969, 1953.

Harris, K., Huntington, D., and Sholes, G., Coarticulation of some disyllabic utterances measured by EMG techniques, Haskins Reports.

Harris, K., Rosov, R., Cooper, F., and Lysaught, G., A multiple suction electrode system, Electroenceph. Clin. Neurophysiol., 17, 698-700, 1964.

Hayes, K. J., Wave analyses of tissue noise and muscle action potentials, J. Applied Physiol., 15, 749-752, 1960.

Hood, R., Some physical concomitants of the perception of speech rhythm of the deaf, Ph.D. Dissertation, Stanford University, 1966.

Hudgins, C.V. and Numbers, F.C., An investigation of the intelligibility of the speech of the deaf, Genet. Psych. Monog., 25, 289-392, 1942.

Huntington, D.A., Harris, K. A. and Sholes, G.N., An electromyographic study of consonant articulation in hearing-impaired and normal speakers, J. Speech Hearing Res., 11, 147-158, 1968.

John, J. E. J. and Howarth, J. N., The effect of time distortions on the intelligibility of deaf children's speech, Language and Speech, 8, 127-134, 1965.

Liberman, A. M., Cooper, F. S., Shanweiler, D. P., and Studdert-Kennedy, M., Why are speech spectrograms hard to read? American Annals of the Deaf, 113, 1968.

Liberman, A., Delattre, P., Cooper, F., and Gerstman, L., The role of C-V transitions in the perception of the stop and nasal consonants, Psychological Monographs, 68, 1-13, 1954.

Liberman, A., Delattre, P., Gerstman, L. and Cooper, F., Tempo of frequency change as a cue for distinguishing classes of speech sounds, J. Exp. Psychol., 52, 127-137, 1956.

Lindblom, B., Accuracy and limitations of Sonagraph measurements, Proceedings of the IVth International Congress of Phonetic Sciences. 1961 - Helsinki, Mouton & Co., The Hauge, 188-202, 1962.

Lindblom, B., Spectrographic study of vowel reduction, J. Acoust. Soc. Amer., 1773-1781, 1963.

Lisker, L., Closure duration and intervocalic voiced-voiceless distinction in English Language, 33, 42-49, 1957.

Lisker, L., Minimal cues for separating /w, r, l, j/ in intervocalic positions, Word, 13, 257-267, 1957.

MacNeilage, P.F., EMG and acoustic study of the production of certain final clusters, J. Acoust. Soc. Amer., 35, 461-463, 1963.

MacNeilage, P.F. and DeClerk, J.L., On the motor control of coarticulation in CVC monosyllables, J. Acoust. Soc. Amer., 45, 1217-1233, 1969.

MacNeilage, P.F. and Sholes, G. N., An electromyographic study of the tongue during vowel production, J. Speech Hearing Res., 7, 209-232, 1964.

Moore, J. C., Fabrication of suction-cup electrodes for electromyography, Electroenceph. Clin. Neurophysiol., 20, 405-406, 1966.

Ohman, S. E. G., Coarticulation in VCV utterances: Spectrographic measurements, J. Acoust. Soc. Amer., 39, 151-168, 1966a.

Ohman, S. E. G., Perception of segments of VCCV utterances, J. Acoust. Soc. Amer., 40, 979-988, 1966b.

Ohman, S. E. G., Studies of articulatory coordination, Quarterly Progress and Status Report, Royal Institute of Technology, Stockholm, Sweden, April, 1967.

Ohman, S. E. G., Speech analysis: Duration of formant transitions, Quarterly Progress and Status Report, Royal Institute of Technology, Stockholm, Sweden, 1, 10-13, 1965.

Stevens, K. N. and House, A. A., Perturbation of vowel articulation by consonantal context: an acoustical study, J. Speech Hearing Res., 6, 111-128, 1963.

3. The Effect of Whispered Speech on the Intelligibility of a Tone Language

a. Introduction

A widely accepted fact among American and European speakers is that people can be understood without any difficulty when they whisper. This fact is not unreasonable if the formant frequencies of the vowels and envelopes as well as the spectra of the fricative and plosive sounds are considered to be information carrying elements of speech (Gjardmann, 1923-25). However, there is no general agreement as to the intelligibility of whispered speech when it is used in certain African and Asian languages which are tone languages.

In tone languages, pitch is used to differentiate the meaning of various lexical items consisting of otherwise identical segmental phonemes. In these languages pitch serves as a basic phonemic characteristic.

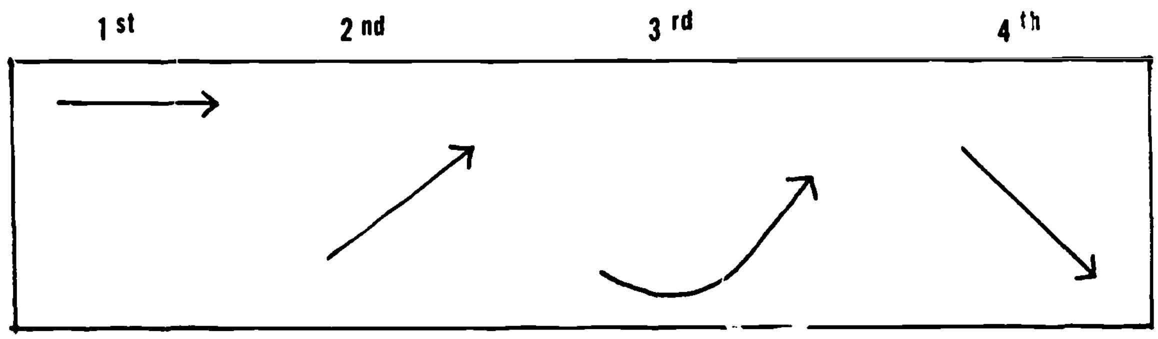
In a true aphonic whisper there is no fundamental frequency or overtone structure because the sound source is not produced by vocal fold vibration. In the absence of a fundamental frequency it would seem impossible, or at least difficult, to distinguish changes of pitch. Nevertheless, if a native Chinese speaker were asked if he had any difficulty understanding a second Chinese speaker who was whispering, the answer would invariably be "no".

The present study will be designed to investigate the effect of whispered speech on the intelligibility of tonal perception in Mandarin Chinese, an important tone language spoken by approximately 500 million people in mainland China, Taiwan, and in various Southeast Asian countries where Chinese immigrants are found. In order to adequately specify the effect of whisper on intelligibility, perceptual judgments and acoustical analyses will be done. One and two syllable words will be used; the two syllable words will be composed of morphemes rather than combinations of words.

b. The Tonal System of Mandarin Chinese

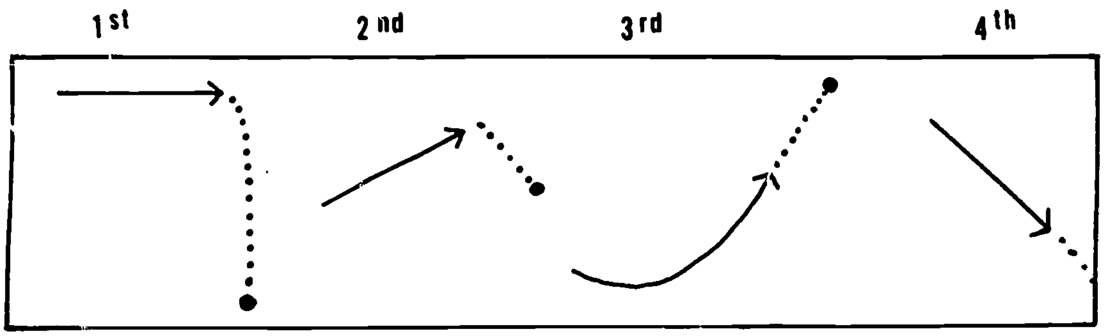
A Chinese word is composed of consonants, vowels, and a constituent tone. Although tonal changes are used in English they are not as distinctive a feature of the language as they are in Chinese. In English, it is possible to say the word "no" in isolation and convey different meanings or emotional moods simply through tonal variations. This is expressive intonation (De Francis, 1963). In Chinese, two words similar in all respects but that of tone will be as dissimilar to a Chinese speaker as "bed" and "bud" are to an English speaker. The tones are learned as part of the syllables with which they are associated. In this integral manner, tones serve to distinguish quite different words as do the vowels /æ/ and /ʌ/ in the English words "bat" and "but". For example, the Chinese "ma" with a steady tone means "mother", while "ma" with a falling-rising tone means "horse".

Mandarin Chinese is composed of four basic tones and a neutral tone. The tones are: 1) high and level; 2) rising; 3) falling-rising; 4) falling; the neutral tone is very short and weak. The tones are contrasting contours of pitch, volume, length, and glottalization (Fischer-Jorgensen, 1961). The four basic tones are illustrated in the following figure in relation to the range of a speaker's voice.



The first tone starts near the upper limits of a speaker's vocal range and continues on that level to the end. The second tone starts at mid-range and rises rapidly to the top of the range. The third tone starts below mid-range, dips to the lowest pitch, and rises above mid-range. The fourth tone starts near the top of the range and falls very rapidly toward the bottom (De Francis, 1963). The actual height and interval of these tones are relative to the sex and voice of the individual, and to the mood of the moment (Chao, 1948).

Stressed syllables always have one of four tones. When the same syllable is unstressed the tone often disappears. Some syllables are never stressed and these unstressed syllables have the neutral tone. If a neutral tone begins an utterance it is pronounced at the mid-range of the speaker's voice and if a neutral tone occurs at the end of an utterance it is affected by the tone of the preceding syllable. The pitch of a neutral syllable occurring after each of the four tones is indicated in the following figure by the large dot.



c. Review of the Literature

In Tone Languages, Kenneth Pike (1948) expresses doubt as to the possibility of changing the pitch of a whisper for a specific vowel unless the vowel is modified somewhat. Pike states that tonemes become ambiguous or undistinguished in whispered speech and that intelligibility depends on context.

Panconcelli-Calzia, in a study cited by Meyer-Eppler (1957) maintains that Chinese born subjects had difficulty understanding whispered Chinese. Panconcelli-Calzia said that whispered speech contains the speech qualities of intensity, duration and sound timbre to a certain degree, but there is no tonal pitch. He also said that in the Chinese confessional there is a loss of intelligibility because of whispered speech.

Franz Giet (1956), who had been a missionary in China for seventeen years, said that in the confessional he had to understand whisper when there were no contextual cues. Giet maintains that one cannot expect educated subjects to make lexical tone judgments from isolated stimulus items in non-whispered speech. Since whisper is normally less intelligible than normal phonation, tonal judgments become more difficult, especially if the listeners are of different dialect backgrounds.

Meyer-Eppler (1957) states that it is not difficult to produce the same whispered vowel on different pitch levels within a range of about a musical fifth. This can be done only by changing the spectral structure of the vowels within limits of recognizability. This is similar to the speculations of Pike, who says that if intensity in whispered speech is not providing the contrast for tonal judgments, the vowel must be modified somewhat. Wise and Chang (1957) suggest that Meyer-Eppler's conclusion that the whispered pitch of a specific vowel can be changed is due to shifting of the dimensions of the supraglottal cavities to those of a different allophone or a different vowel.

Meyer-Eppler (1957) made spectrograms of five German vowels /i/, /e/, /o/, /u/, and /a/. He found formant 3 of /ɑ/ shifted from approximately 2500 cps to 3000 cps when a higher pitch was intended. A similar shift was found at 5000 cps for a weak formant 5. With /u/, formant 1 was raised from 600 cps to 700 cps when a higher pitch was intended. Unfortunately, Meyer-Eppler does not furnish any information as to his speaker and the instructions given the speaker, so it is not possible to determine whether these formant changes were the result of an acutal pitch change or whether they were due to an allophonic variation of the vowels.

There is no unanimity of opinion in the literature as to what acoustic cues enable pitch to be perceived in whispered speech. Meyer-Eppler believes it is the movement or displacement of formant regions which follow the melody "course" that aids pitch perception. Giet believes it is the color differentiation of the vocalic character caused by the lifting and lowering of the larynx for the high and low tones respectively, and by increasing air pressure for the high tones and decreasing air pressure for the low tones (1956, p. 375). Giet uses the word "wortmelos" which has been translated as "word melody" (melos being the Greek for melody or tune), when he talks of the effect of air vibrations on the impressions of "wortmelos" (p. 376). Although Wise and Chang did not perform any acoustic analyses with their data, they believe that as long as the whispered vowel is repeated with no attempts to change its quality as a vowel, the formants will be unvaryingly reproduced regardless of the intended pitch change.

d. Experimental Design

The stimulus items, to be spoken by native-born speakers of Mandarin Chinese, will consist of forty words embedded in the middle of a carrier sentence which will furnish no cues as to meaning. Twenty-one-syllable and twenty-two-syllable words will be chosen from a list of 2000 most commonly used words in spoken Chinese. Each group of words will be composed of four stops, four fricatives, four affricatives, four nasals and four glides. Each of the twenty-one-syllable words will have four meanings and each group of four will be homonyms when uttered without their attendant tones. Of the twenty-two-syllable words, ten will have no alternate meaning, seven will have two alternate meanings and three will have three alternate meanings.

Standard recording procedures will be followed. All stimulus items will be recorded under three conditions for comparative purposes. These are: (1) normal phonation, (2) whisper and (3) monotone whisper. A panel of native born speakers of Mandarin Chinese will act as judges using a series of closed sets of varying size, i.e., number of foils. Based on signal detection theory, the different closed sets will give an estimate of judgmental criteria and sensitivity. Stimulus items will be randomly distributed between the sets. Correctly identified whispered items will be analyzed acoustically for changes in duration, intensity, spectral distribution and formant frequency shifts. The identical measurements will be performed on the same items produced with normal phonation and by a monotone whisper.

The significance of this study is the further specification of the acoustic cues for perception of speech under various conditions. In this case it is normal phonation versus whisper with a language containing tonemes. Since I teach a course entitled Acoustics of Speech I will be able to gain valuable information on different and exotic aspects (tones) of the production of an acoustic signal and its perception.

References

Abramson, A. S., The vowels and tones of standard Thai: acoustical measurements and experiments, International Journal of American Linguistics, part III, 28, 1962.

Chan Shau Wing, Elementary Chinese, Stanford, California, Stanford University Press, 1948.

Chao, Yuen Rin, Mandarin Primer, Cambridge, Massachusetts: Harvard University Press, 1948.

De Francis, J., Beginning Chinese, New Haven, Connecticut: Yale University Press, 1963.

Fischer-Jorgensen, Eli, New techniques of acoustic phonetics, in Psycholinguistics, Saporta (Ed.), New York: Holt, Rinehart and Winston, 1961.

Hockett, C. F., Peking phonology, in Readings in Linguistics, M. Joos (Ed.), New York: American Council of Learned Societies, 1963.

Giet, F., Kann man in einer tonsprache flustern? Lingua, 5, 372-381, 1956.

Gjerdmann, O., Critical remarks on intonation research, London University, School of Oriental and African Studies: Bulletin, 3, 495-505, 1923-25.

Luchsinger, R. and Arnold, G., Voice-Speech-Language, Belmont, California: Wadsworth Publishing Co., 1965.

Meyer-Eppler, W., Realization of prosodic features in whispered speech, J. Acoust. Soc. Amer., 29, 104-106, 1957.

Pike, K., Tone Languages, Ann Arbor, Michigan: Edwards Brothers, 1948.

Tsai, Le-Sheng, Chang Yung Tze Hsiian (Most Frequently Used Chinese Words), 1946.

Wise, C. M. and Chang, L. P. H., Intelligibility of whispering in a tone language, J. Speech Hearing Dis., 22, 335-338, 1957.

II. Interdisciplinary Research

1. Research in Diver Speech and Communication (with Professor Hollien)

a. Introduction

This section describes an ongoing program of research supported by the Office of Naval Research. It is included in order to provide a greater perspective of the overall research program of the Communication Sciences Laboratory. The program has two major thrusts, each one divided into two sections. They are as follows:

(1) Basic Research

(a) in underwater speech production and perception

(b) in underwater auditory acuity and function*

(2) Applied Research

(a) in diver communication; in evaluating and testing diver communication systems and helium/oxygen unscramblers

(b) in underwater auditory acuity and sound localization*

Man is increasingly turning to the sea for scientific, military, economic, ecologic and recreational purposes. The two approaches used for such explorations are underwater submersibles and the "free" diver in the sea. Often these two methods are combined but, whatever the purpose of an underwater expedition, its effectiveness and success depends, to a great measure on the effectiveness of divers as workers in the sea. Yet, at present, divers are not efficient workers in this milieu because of a lack of adequate voice communication between individuals comprising a diving team, between diving teams and/or surface support personnel (i.e., by submerging they acquire immediate and severe speech and hearing disorders). Moreover, the need for voice communication is particularly urgent: 1) where visibility is limited; 2) where even relatively minor equipment malfunctions can prove fatal unless assistance is immediately available; and 3) where dangerous marine life or work situations are existent.

While many research papers and symposia have been addressed to questions related to marine bioacoustics, little is known about man's own ability to communicate while underwater. To do this, the unique difficulites imposed upon communication by a liquid environment must be considered. In water, gestures and facial expressions, which are habitual cues in air, dwindle markedly due to

---

* For details of the research program on underwater acuity and sound localization refer to the section by Hollien and Feinstein.

restricted visual parameters including the use of a number of life-support devices (regulators, masks, and so forth). Writing is limited, awkward and slow as a method of underwater communication as are systems such as Morse Code. Therefore, if normally-paced communication is to take place underwater, speech must carry the main burden. However, speaking directly into the water presently is impossible; additionally, recognition and interpretation of speech signals transmitted through the water is sharply limited by the sensitivity of the submerged human ear.

The Communication Sciences Laboratory has developed and has been carrying out a program designed to answer fundamental questions about diver communication and to acquire basic knowledge about the factors that limit or allow man to communicate underwater. This program focuses on several areas of investigation: 1) studies of man's ability to produce intelligible speech under the constraints he encounters as a diver, 2) studies of underwater speech propagation and the various effects of bottom-surface and thermocline wave guide channels, distance, filtering, masking and other distorting underwater characteristics on speech intelligibility, 3) the analysis and appraisal of diver's voice communication systems and habitat communication systems, and 4) development of specialized instrumentation that will permit conducting underwater research with precision similar to that available in speech communication research conducted in air.

In order to carry out this program of research, the Communication Sciences Laboratory has trained a number of faculty and graduate students with expertise in experimental phonetics and acoustic phonetics, psychoacoustics and electrical engineering to be scientific divers. Our group is probably the only group in the United States that is working together on an integrated program that is dedicated to a comprehensive study of basic and applied research in underwater communications.

The following section will present some of the ongoing studies in diver speech communication and systems evaluation.

b. Development of a Diver Communication System (DICORS)

It is obvious that if precise and rigorous underwater research is to be carried out successfully, the investigational procedures cannot be haphazard. For example, research on speech communication in air involves a substantial variety of highly sophisticated techniques and methodologies and these approaches permit a precision and rigor in that milieu presently unavailable underwater. In an attempt to utilize such methodologies and minimize as many of the extraneous variables as possible, we have developed an underwater system which provides for experimental control of diver positioning, stimulus presentation, and subject response. The total equipment configuration has been named "Diver Communication Research System" (DICORS).

The overall configuration of DICORS is that of an open framework in the shape of a truncated prism standing on one end. Its dimensions are as follows: height--80 inches; depth--34 inches; width at the back (at diver's seat)--46 inches;

and width at the front--22 inches. In addition, DICORS has a 22-inch frontal extension which provides mounting for certain items of equipment. The framework of the system is constructed of poly-vinyl chloride (PVC) tubing. The main frame consists of 1.5 inch ID schedule 40 PVC tubing, the cross braces are of 3/4 inch schedule 80 PVC tubing. The framework is free flooding with all potential cavities provided with air escape and water drain holes. Lead anchors, which have been attached to the bottom of each of the four main (vertical) structural members, provide adequate negative bouyancy to allow the unit to hang stably in the water from its sling and suspension cable. The main support of DICORS is provided by nylon ropes attached to the sling, pass through the main vertical members and secured to eye bolts within the lead anchors. In order to provide further stability and to prevent rotation of the unit on its axis, two guy wires were passed through eye bolts attached to the top and bottom of the two rear vertical members. These guy wires also provide safety stops in case of the accidental release of the sling or suspension cable.

In order to control the diver's position with respect to research equipment, he is situated on the seat with his feet positioned either on the cross-members or hanging free. His head is placed in a positioner-(three types of head positioners are used depending on the nature of the research being conducted)--and a weight belt, placed across his lap, assists in holding him in the proper position. Under these conditions, not only is the diver positioned properly for the particular research procedures being employed, but he can also return readily to the same position for replications of the procedure or for other projects. Moreover, ingress and egress to DICORS is quick and simple.

In addition to the DICORS described above we have developed several mini-DICORS. The mini-DICORS was designed for portability as well as for use as part of a work study to be described below.

c. Speech Studies

(1) Standardization of Speech Materials for Underwater Communication Research

A rather substantial series of studies are in progress in which currently available materials for use in assessment of underwater speech intelligibility are being evaluated. The research approaches are based on experiments completed at Bolt, Beranek and Newman -- and on an approach reported by Williams at the November, 1967 ASA convention. The word lists provided by Black, Fairbanks, House, Voiers, Griffiths, Clarke and others, are under investigation; as are the Speaks/Jerger synthetic sentences. A CSL "Profile" test (composed of vowels, monosyllable words, rhyming words and synthetic sentences) has been prepared for research and for use in system evaluations of all types.

Basically, this series of projects has grown in scope to include both standardizations of speech materials and the study of needed and preferred words (and phrases) as judged important by both military and civilian divers. This latter aspect is developing into a Diver Lexicon comprised of messages with relevancy to such situations as work-in-the-sea, safety-emergency, habitat operation and so on.

(2) Error Analysis of Divers' Speech

A major series of projects was initiated last year. These studies have focused on the phonemic analysis of divers' speech and the errors typically made under conditions of high ambient pressure, exotic gas mixtures, underwater communication equipment and so on. Previously, very little research has been carried out which had attempted to analyze the types of errors that these varied speaking situations induce upon the speech of the diver. Since it is most important to discover what specific types are introduced by the distorting effects of the various constraints listed above, initial studies attempted to identify phonetic classes that are most affected by 1) high ambient pressure, 2) helium-oxygen breathing mixtures, 3) adding a cavity to the vocal tract, 4) the back pressure of the SCUBA breathing apparatus and so on. In all cases, attempts were made to simulate, in the laboratory, the environmental characteristics under examination - or to use existing recorded speech materials. However, as a second phase of this project, additional data will be gathered and analyzed with the diver in the actual underwater milieu associated with that phase of the research.

(3) Speech in a Helium-Oxygen (HeO_2) Environment

This area focuses on investigations of speed distortions caused by breathing mixtures of HeO_2 under high ambient pressures. The emphasis in these studies is placed on speech intelligibility measurements and on the puzzling, non-linear vowel formant shifts. Other investigations in this area are concerned with diver's speech production and reception adaptation, particularly with respect to their apparent ability to eventually improve speech intelligibility by experimenting with articulation. It is important to discover just what speech sounds are most affected by the milieu, as well as obtaining an index of a general speech intelligibility level. It has been determined that speech intelligibility is reduced as a result of depth and the HeO_2 mixture (at 600 feet for example, speech intelligibility levels are around 20%). Subjects for most of these studies are Navy aquanauts when working in underwater habitats in a saturated condition.

As part of our basic research on methods to diminish the speech distorting effects of the HeO_2/high ambient pressure environment, twelve Communication Sciences Laboratory divers descended to 300' in an HeO_2 mixture at the Westinghouse Ocean Research and Engineering Center's hyperbaric facility. All read Griffiths word lists under the following conditions (the order was counterbalanced): 1) normal, 2) high f_o, 3) low f_o, 4) fast speech, 5) slow speech, 6) high vocal intensity, and 7) most intelligible (as judged by talker); one half of the divers had no side-tone as auditory feedback was masked out by an 85 dB masking signal; analysis of the materials is in progress. It is expected that some data will result that indicate what speech factors (among those studied) will serve to enhance intelligibility.

In addition to studying the speech of the diver in HeO_2 we are also conducting a four-phase evaluation of helium unscramblers which is designed 1) to determine the exact nature of the equipment; 2) to develop a standard test for evaluating all types of unscramblers; 3) to test unscramblers on-line 4) to test unscramblers off-line. A short review follows.

A series of investigations evaluating HeO_2 speech unscramblers produced by a) the Naval Applied Sciences Laboratory, b) the HRB-Singer Co., c) General Precision/Singer, d) the Westinghouse Corp., e) the Raytheon Corp. and f) Industrial Research Products, Inc. have been conducted. In the first study, four divers produced PB_{25} word lists at EDU at a simulated depth of 600 feet. Recordings were made on a Honeywell 8100 FM tape recorder which simultaneously recorded the unprocessed speech and the output of the three unscramblers used on-line. The first procedure involved evaluation of intelligibility levels recorded immediately when the divers reached the experimental level of 600 feet. Subsequent evaluations were conducted wherein the variables are a) changes in diver intelligibility over time, b) comparison of the Electrovoice and Roanwell microphones, c) comparison of various MDL microphones and d) comparison of the MDL and Scott masks used underwater. The NASL and Westinghouse units showed promise in these tests.

An off-line evaluation of the NASL, Westinghouse and IRPI unscramblers also is now complete. In order to carry out this type of research, a special test had to be developed. Stimulus tapes included speech samples 1) produced by a number ot talkers, 2) at a number of simulated depths, 3) at a number of HeO_2 mixtures, 4) as a function of time, and 5) as a function of intelligibility level and so forth. It was found that the NASL and IRPI unscramblers increased intelligibility to overall levels of about 35-40%; substantial improvement but not adequate for good voice communications in this (or any other) environment. In the most recent evaluation, the IRPI, Raytheon and General Precision/Singer units were evaluated on-line at the Westinghouse Hyperbaric facility, Annapolis, Maryland. All units were operated in conjunction with five microphones; depth was 650 feet; talkers were three divers. The overall intelligibility levels were 1) unprocessed: 15.7%, 2) IRPI: 32.7%, 3) G-P/Singer: 21.5%, and Raytheon: 45.1%. Conclusions were that the Raytheon exhibited the best performance followed by an improved IRPI unit and none of these devices yet allow for adequate communication. A new set of investigations is planned; the research will be replicated until a reasonable solution is forth coming.

(4) Studies in Underwater Speech Propagation

Included in the research on the transmission of speech through water are studies of the effects of distance, filtering, and noise masking on speech signals. An example of this type of investigation is one in which speech intelligibility was functionally related to distance. The principal interest here was the phase distortion effect of the ocean surface and bottom (acting as a wave guide) on the signals being transmitted through the fluid medium. It was found that the major degradation in speech intelligibility results not from phase distortion but from the masking effect of ambient underwater noises. In another study, filtered speech was evaluated both in conditions of quiet and noise, the results indicated that intelligible speech can be transmitted underwater and that propagation of such signals obey perceptual rules similar to those in air.

(5) Intelligibility of Diver-to-Surface Communication Systems as a Function of Distance

While impressive gains have been made in basic equipment for divers, systems for voice communication are still relatively primitive. Nonetheless, a number of underwater speech communication systems, both civilian and military, have become available to divers. However, so little is known about man's basic ability to speak underwater that the design of these systems has been, of necessity, based primarily upon electronic considerations. Moreover, few systematic or independent evaluations have been carried out on these units. The need remains, then, for an assessment of system efficiency in the transmission of intelligible speech under conditions designed to duplicate actual diver-to-listener communication. The current project is the third in a series of six planned evaluations we are carrying out on such systems. They are: Diver-to-surface 1) at close range in fresh water, 2) off-shore in saltwater -- as a function of range (the present study), and 3) in a saltwater harbor -- as a function of range. Diver-to-diver communication is also evaluated under these same sets of conditions.

The procedures used to gather data for the current diver-to-surface study (over distance) essentially parallel those previously used by our group. Basic to the standardization of such underwater procedures is our Diver Communication Research System (DICORS) which has been described above. For the present diver-to-surface study conducted at Buck Island, the Virgin Islands during TEKTITE-2, we used the small and portable mini-DICORS.

The mini-DICORS was floated at about 15 feet in approximately 30 feet of water on the Buck Island range. Hydrophones were situated at distances of 50, 250, 500, 1000, 2000, and 4000 feet from the diver/talker. Since depth varied as a function of distance down-range, the hydrophone was placed halfway between the bottom and surface at each location.

Two basic types of communication systems were evaluated; the first group consisted of acoustic systems. An "acoustic" system includes a microphone, amplifier, power supply and transducer; it characteristically transduces speech directly into the water by means of the projector (underwater loudspeaker). The signal produced can be received by a hydrophone placed in the water, or by divers without any special receiving equipment. Two of the acoustic systems studied were the Raytheon Yack-Yack and the Bendix Watercom, (we had modified the Watercom to enhance its operation); both were evaluated in a configuration that included a double hose regulator and Bio-engionics (Nautilus) muzzle. A third "acoustic" system evaluated was the Scuba-com; it is a mechanical (rather than electronic) unit that consists of a small air filled cavity and diaphragm.

The second group of communicators consisted of amplitude modulated (AM) systems and included the Aquasonics 420, the ERUS-2-3A, the SubCom prototype (both short range and intermediate range) and the military PQC-2. All systems evaluated were rigged with the Bio-engionics "Nautilus" muzzle and a double hose regulator. In an AM system, a carrier wave is utilized and modulated by the speech signal. Such a system ordinarily consists of a microphone, power supply, amplifier, modulator and underwater transducer. Speech produced in this manner can be understood only by a diver or a surface observer having an appropriate receiver and demodulator.

With the Aquasonics unit, a 42 kHz carrier signal is transduced into the water after being modulated by the speech signal and the mixed signal is picked up by a receiving coil, demodulated, and heard in the normal speech mode. The PQC-2 (a military unit) uses a SSB suppressed carrier frequency signal of 8.0875 kHz; the ERUS-2-A, a French developed unit, also utilizes an acoustic carrier of 8.0875.

Obviously, the purpose of this study was to provide information on how the units listed above would perform over distance and, unlike our previous fresh water studies, how they would perform in a saltwater environment with its attendant sea noises. Each diver/talker (N=4) was assigned a 50-word Griffith list (equated for difficulty) to be read with each communicator at each distance. Each word was preceded by the phrase, "You will say--".

In order to evaluate speech intelligibility or the intelligibility of communication systems, tape recordings of the lists read by the divers are brought to the University of Florida and played to a minimum of ten "semi-trained" listeners i.e., students selected on the basis of 1) being native speakers of English, 2) having normal hearing and 3) being capable of performing the required listening tasks. Before hearing the tapes, listeners are required to score at least 92% on a screening test which included 50 words from CID Auditory Word List A-3 (Hirsh recording) recorded in +10 dB of thermal noise, 25 words recorded in a HeO_2 environment, 25 words from the communicator recordings, and 50 words from CID Auditory Word List 4-A. The final 50 words constituted the screening test. This study is in the data reduction phase; additional studies (see above) are planned over the next 3-5 years.

(6) Effectiveness of Work by Divers With and Without Voice Communication

Finally, it would seem appropriate to illustrate another of our research programs by the following initial study. Man's tasks under the sea have become more complex and major obstacles to the successful accomplishment of work requiring a cooperative endeavor among members of a diving team still exist, i.e., the lack of communication among team members and/or their inability to effectively use the communication equipment available. In an effort to study and quantify the problems inherent in an underwater work situation, a pilot study involving two teams of divers engaged in an identical work situation was completed at TEKTITE-2, Virgin Islands, May, 1970. This investigation, which alternated matched teams, was designed to determine whether a team of four divers with communication could accomplish a complex construction task (requiring a high degree of cooperation among them) more effectively than could a similar team without communication. Two four-member teams, matched in ability but unused to working with each other and unused to using communication systems, served as subjects. The work task consisted of assembling a mini-DICORS (132 parts) in 20 feet of low visibility water. On the first trial, members of Team A wore aquasonics 420 units; Team B had no communication gear. On the second trial (a day later), the communicators were worn by Team B. Time of assembly, number and type of errors constituted the objective measures. In all cases, the performance of the team without the communication gear was superior. It was concluded that to be aided by communication systems 1) divers must be trained in their use, 2) they must be allowed to develop appropriate communication procedures and 3) systems with better intelligibility than those currently available must be obtainable. As a continuation of this study, we plan to conduct two types of research; one similar in nature to the TEKTITE-2 research (i.e., analogous tasks for matched teams with and without communication) but after extensive diver training is completed. In the second study, the effects of communication use after training will be investigated with divers who are carrying out scientific and exploratory tasks in various underwater programs throughout the state of Florida.

Stanley Y.W. Su

I. Personal Research

1. Data-Sharing and Man-machine Interaction in Information Systems

a. Introduction

As a result of scientific and technological progress, files and volumes of information in various forms have been proliferated in our society. Many people and organizations have become concerned with the problem of processing data and utilizing relevant data to make intelligent decisions. Modern digital computers have been used for this purpose and have been found to be effective and useful for this task due to their high speed, their precision and their vast storage capacity. Many information systems have been built to handle data of various forms. In general, each uses a data base having a data structure and organization different from the others. Information processed in one system cannot be shared by the other systems. This situation causes unnecessary duplication of efforts in data collection and information processing, as well as in system implementation. To avoid unnecessary duplication, we need to form a network among information systems so each can have access to the data for other systems.

Recent progress in equipment technology has introduced many sophisticated storage media, remote terminals and optical devices, and has made possible the development of information systems for the analysis, recognition, storage, retrieval and display of data for various applications. The development of on-line information systems has brought the user one step closer to his data and has necessitated the study of the problems of man-machine communication.

This proposal deals with the study of some basic problems dealing with data-sharing among information systems and with ambiguity resolution and feedback utilization in man-machine communication.

b. The Proposed Research

The following paragraphs describe the research problems to be investigated and the approaches to be taken.

(1) Data Sharing in Network Systems

In the past few years, research effort expended on information storage and retrieval has been very intensive and broad in scope. The use of large digital computer systems allows large quantities of scientific data to be processed and stored in the computer memory and information to be retrieved to answer various types of information requests. There are many operational systems implemented by the government, universities and commerical companies. Some examples are the systems in the Library of Congress and the Defense Documentation Center, the MEDLARS, the SMART system, IBM's GIS and PARS, CDC's INFOL, and Auerbach's DM-1. The existing systems are operated on different data bases with different data structures and file organizations. They are in general implemented on different machines with different hardware equipment and software facilities. Consequently, the data collected and processed by one system generally cannot be used by the other systems. This problem of not being able to share data among different applications in one system and among different information systems has resulted in a duplication of effort and has prevented the development of an interaction among information systems.

One solution to this problem would be to force all information systems to store their new data in an identical or compatible form, and reprocess the old data into this new commonly acceptable form. However, the reprocessing of old data will be too costly even if a general and compatible data structure can be designed. Moreover, data stored in such a structure may physically be stored in different secondary storages in different systems. One system would still not be able to have direct access to the data base of the other systems. A more realistic approach would be to convert the data stored in one structure by one system to fit another data structure required by a second system whenever data-sharing is requested. This second approach would allow all systems to continue using the structures which are most suitable for their own applications. Only the data requested by other systems would need to be converted.

The network project sponsored by the Advanced Research Project Agency (ARPA network) is designed for resource sharing, i.e. sharing of hardware and software facilities as well as data sharing among systems in the network. Until recently, the project has concentrated on the problem of physically connecting the hardwares together as reported by Roberts and Wessler (1970), Heart et. al. (1970), Frank et. al. (1970), Carr, et. al. (1970) and Kleinrock (1970). The software specifications and requirements for the application of the network have not been fully studied.

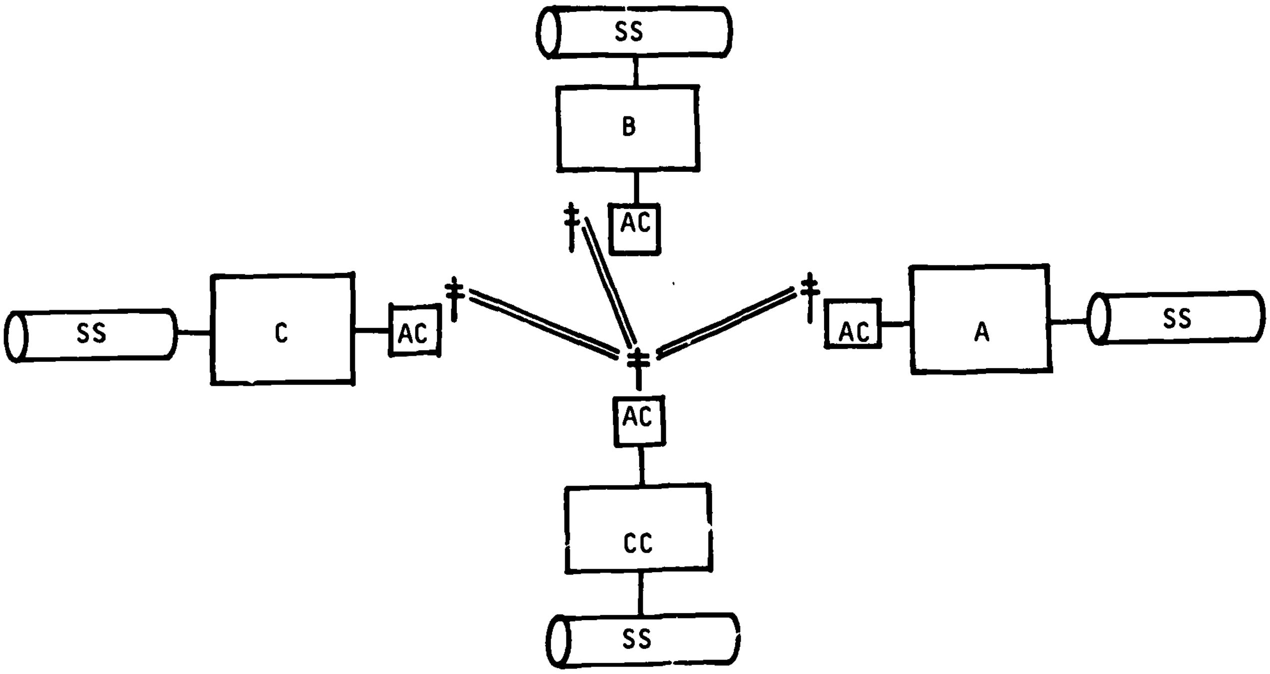
In order to utilize a computer network to share scientific data among information centers, the following software tasks must be carried out:

(a) the design of a data manipulation language to allow the user of the network to specify the records, files and data sets he wants from the other systems.

(b) the design of a data description language to formally define the data structures of all the data bases so that the network control program can convert data from one structure to another.

(c) the specification of the relationship between the network control program and the operating system of each individual system in the network. The operation of the whole network should not jeopardize or drastically reduce the efficiency of any system to perform its local computation tasks.

The hardware configuration of the proposed network system for data-sharing is the STAR configuration. In this network configuration, all the information systems denoted in the figure as A, B, and C are connected to the Communication Center (CC) via public or leased telephone communication lines. Each system has an acoustic coupler (AC) to receive and transmit data stored on a secondary storage (SS).



In the proposed system, a data manipulation language (DML) will be formulized to allow the users of all systems to state their requests. This language will be of the same type used in some existing data management systems as the language used in the IBM's GIS. It allows the user to specify the names of data items, records, sets of records or files in which data are contained and also the search criteria specifying the type of data for which the user is looking. The names specified have to be the ones used in the requested system. Therefore, a data description of one system should be made available to the users of other systems. The statements in this language will be transmitted to the Communication Center. The Communication Center serves as an interpreter for the various systems in the network which may have different hardware facilities, may use different application programs written in different programming languages and may request data stored in other systems in different structures. The main functions of the Communication Center would be as follows:

(a) It will 1) receive data requests posed in the DML from systems requesting data, 2) interpret the requests and 3) establish a connection with the proper systems from which data is requested.

(b) It will transmit to the requested systems the names of the records, sets or files requested.

(c) It will receive data transmitted from the requested systems and select those data which satisfy the search criteria specified by the requesting systems. The selected data will be converted to fit the data structures of the requesting system.

(d) It will send the data to the requesting systems in formats suitable for their applications.

In the proposed system, each information system will have a program to interact with the Communication Center. The program will accept a list of data names transmitted from the Communication Center. It will determine which data file contains the requested data, and call a library program, written specifically for that file, to obtain data from secondary storage. Each library program can be written in the same programming language as that one used for writing the program which generated the file. The programmer who produced the original file would have no problem in writing such a program to read data from the secondary storage into the core storage. Data associated with the received data names will be sent to the Communication Center for further analysis and conversion. I believe that this approach would solve the problem of hardware and software incompatibility among information systems.

The initial step in this research would be to write a simulation program to test the data-sharing operations described above. The simulated system will involve three subsystems: two information systems and the Communication Center. The simulation program will be written in PL/1 and will use the multitasking facility in the language to allow various tasks (the simulated systems) to be processed concurrently. The simulated system can then be expanded into a full scale network system in which the Communication Center is a time-sharing system capable of servicing the data-sharing requests of all systems in the network concurrently. The IBM 360 model 65 will be used to implement the initial simulation system as well as the final time-shared communication system.

(2) Man-Machine Interactions

Although batch processing is adequate for many computer applications such as payroll processing and report generation, an interactive computer system using on-line display facilities would allow researchers to enter, edit, analyze and retrieve data in real time. By using an on-line interactive system as a means of readily obtaining and analyzing his data, a researcher will be stimulated to form new ideas and techniques to carry out his work.

Many technical problems dealing with file maintenance, on-line editing, data analysis and optical scanning have been separately tackled by many display systems such as MIT's MAP, the Culler-Fried system, the Video-Display system in Medical Research (Worley 1969), NASA's AMTRAN, SDC's DISPLAY and Harvard's TOC. My major emphasis with regard to man-machine interaction will be on the study (1) of how to make the best use of the user's innate abilities to tolerate ambiguities and to account for context in making judgments and (2) of the machine's speed, precision and vast memory, to accomplish the communication tasks demanded by the user. More precisely, the following two problems will be investigated.

(a) User's and System's Roll in Ambiguity Resolution

From the user's standpoint, the best language for man-machine communication is the language he uses daily, i.e., a natural language. However, natural language abounds in ambiguities of various kinds. The problem of ambiguity is vital to any information system using natural language as its communication language. Many natural language information processing systems have been implemented. In a recent survey, Simmons (1970) pointed out the fact that the method of using the stored information in a data base to resolve ambiguities is hardly touched upon.

In an interactive system, the user can be requested to help the system resolve the ambiguities in the input information and search requests. However, in order not to overburden the user, a good information system should do all it can before turning to the user for help.

In analyzing an input text or search request, some word ambiguities can often be resolved by the syntactic parser by taking into account the structural context in which the words occur. Syntactic ambiguities can often be resolved on the basis of semantic information of the words in the input and also words in the other sentences of the same discourse. The objective of my investigation with regards to ambiguity resolution is to study the ways in which the semantic analyzer can use the semantic information in the data base to resolve ambiguities in the input. Naturally, it will be too costly to search the whole data base for each ambiguity encountered in the input. One approach to this problem is to establish a discourse file each time a user is entering information or requesting information. This discourse file will contain part of the information entered by the user and his previous requests. The interpretation of the subsequent input which is semantically related to the information contents of the file will be regarded as a more appropriate one. For example, if the information entered in the system by the user is in regard to the results of a game, the discourse file may contain those high frequency words such as 'game', 'score', 'player', 'contest', 'win', 'lose', etc. These words can be used, for example, to resolve the ambiguity of the sentence "The man hit the pitcher with a piece of rock" despite the ambiguous interpretations of 'pitcher' as a player and 'pitcher' as a container. The selection of 'pitcher' as a player as the proper meaning would be based on the fact that the semantic distance between 'pitcher' as a player to the words in the discourse file is less than the distance between 'pitcher' as a container and the words in the file. Works on the measurement of semantic distance between words based on their distribution are reported in Su (1968, 1969). This approach is taken under the assumptions that (1) the user tends to request the same or related information, (2) words, though ambiguous, are often used by an individual to mean certain things. Once an ambiguous word is resolved either by the system alone or with the help of the user, its proper meaning should be retained so that when the same word is used by the same user, the most probable meaning of the word can be determined. The structure and organization of discourse files and their relation to the main data base will be investigated.

(b) System's Query and User's Feedback

More often than not, the user's initial query to the system contains far less information than the amount which the user has the potential to provide. I believe that it is the system's responsibility to help the user by refreshing his memory so that he can provide the system with more information concerning his needs. The user should play the role of a decision maker at every stage of the retrieval process. The specific research problem I would like to investigate is the practicality of programming the system to generate specific questions for the user and to accept the user's answers as feedback information to optimize the next retrieval operation. I shall elaborate this in the following paragraphs.

The approach under investigation is to have the system expand the user's query and at the same time reduce the document space by using the user's answer to some system-generated questions. The following example will illustrate the approach. Assume that a user has a specific mathematics book in mind but does not remember the precise descriptive information. He may request the system to help him find the book. Initially, the system will search data in the data base for all mathematics books. Instead of outputting the titles of all the books related to mathematics, the system checks the descriptive contents of each candidate in the list. Suppose D_1 is a mathematics book published in the United States in 1940 and written by John Smith dealing specifically with algebra. The system will output the following questions in the form of attribute-value pairs at a terminal.

Output Questions	User's Answer
Place of publication: United States	YES
Author: John Smith	DON'T KNOW
Subject Matter: Algebra	NO
Date of publication: 1940	DON'T KNOW

The user may answer these simple questions and the system uses the answers to eliminate the candidates in the original list. If there is more than one candidate in the list, the system will again check the descriptive contents of a book, say D_2, which is a mathematics book written by Robert Jones, published in the United States in 1960 and dealing with calculus. This time the system will output only the question 'subject matter: calculus'. The questions 'author: R. Jones' and 'date of publication: 1960' will not be asked since the previous questions concerning author and date generate the answer 'don't know'. This question-and-answering process can continue until the proper book is found. Through the feedback loop, the initial query is continually expanded whereas the document space (the initially retrieved documents) is gradually reduced.

c. Significance of the Proposed Research

The proposed research relates to communication science in two ways: the communication among groups of people via information systems and the communication among individuals using the modern digital computer as a medium. The proposed research on data-sharing network systems deals with the establishment of a communication network among information centers to transmit research data or any type of information from one system to another. The proposed system, when implemented, will greatly increase the current awareness of all people who have access to any of the information centers in the network. This will eliminate unnecessary duplication of efforts in all human endeavors and will help individuals make intelligent decisions.

The proposed research on ambiguity resolution is essentially a study of the syntactic and semantic problems in languages. In order to build or program a machine to resolve ambiguities in a natural language, one has to feed the machine with the information concerning the syntax and semantics of the language used. Thus, programming a machine to converse with a man in a natural language provides a good testing ground for testing and evaluating the linguistic assumptions we make concerning the properties of the language. These assumptions are those rules which we build into the same machine. Therefore, this proposed study will contribute to the theoretical study of the syntax and semantics of natural languages. Moreover, the success of this study would bring us one step closer to the idea of man using his own language to present to the machine computational problems without the use of a programming language. The research problem of feedback utilization relates to the general problem of information storage and retrieval. This study will contribute to our understanding on how to make use of information provided by the user to improve the machine's performance.

References

Becker, J. and Olsen, W.C., 'Information Networks', Annual Review of Information Science and Technology, Vol. III, Cuadra (Ed.), Britannica, 1968.

Carr, C.S., Crocker, S.D. and Cerf, V. G., 'Host - Host Communication Protocol in the ARPA Network', Proceedings of the Spring Joint Computer Conference, 1970.

Frank, H. Frisch, I. T. and Chou, W., 'Topological Considerations in the Design of the ARPA Computer Network', Proceedings of the Spring Joint Computer Conference, 1970.

Heart, F. E., Kahn, R. E., Ornstein, S. M., Crowther, W. R. and Walden, D. C., 'The Interface Message Processor for the ARPA Computer Network', Proceedings of the Spring Joint Computer Conference, 1970.

Kay, M. and Su, Stanley Y. W., 'The Mind System: The Structure of the Semantic File', RM-6265/3 - PR, the Rand Corporation, Santa Monica, California, 1970.

Kleinrock, L., 'Analytic and Simulation Methods in Computer Network Design', Proceedings of the Spring Joint Computer Conference, 1970.

Overhope, C. F. J., 'Information Networks', Annual Review of Information Science and Technology, Vol IV, Cuadra (Ed.), Britannica, 1969.

Roberts, Lawrence G. and Wessler, Barry D., 'Computer Network Development to Achieve Resource Sharing', Proceedings of the Spring Joint Computer Conference, 1970.

Simmons, R. F., 'Natural Language Question - Answering Systems: 1969', Communication of the Association for Computing Machinery, Vol. 13, no. 1, January, 1970.

Su, Stanley Y. W., 'A Semantic Theory Based Upon Interactive Meaning', Computer Science Technical Report #68, University of Wisconsin, 1969.

Su, Stanley Y. W. and Harper K., 'A directed random paragraph generator', Preprint No. 13, International Conference on Computational Linguistics, 1969.

Su, Stanley Y. W., 'Managing semantic data in an associative net', Proceedings of the Sumposium on Information Storage and Retrieval, University of Maryland, 1971.

II. Interdisciplinary Research

1. A Discourse Analysis System and Its Application to Computer-assisted Instruction (with Robert L. Moore)

Most of the existing works on natural language information processing have been restricted to the analysis of isolated sentences. Consequently, much of the semantic information contained in textual materials cannot be recognized by the systems. The lack of knowledge concerning the formal properties of language elements beyond sentence boundaries has hindered progress in many areas of studies in information science.

This work attempts to apply the theoretical concept and the knowledge gained from experiments with a paragraph generation system (Su and Harper, 1969; Su, 1971) to the analysis of sentences in a connected discourse. The analysis system will have direct applications in computer-assisted instruction, content analysis, and information storage and retrieval.

In the paragraph generation system, the paragraph is formally described in terms of the attributes of "development" and "cohesion". Many inter-sentence patterns and rules of cohesion in paragraphs of a real discourse have been found and tested in the generation system. These patterns and rules, along with others to be verified in future experiments, will be used in the analysis system to determine the structure and semantic content of sentences in a connected discourse.

The immediate application of the analysis system will be the development of a powerful computer-assisted instruction system for teaching students the basic concepts of journalistic writing. This application is an extension of an existing CAI system.

References

Simmons, R. F., "Natural Language Question-Answering Systems - 1969", Comm. ACM, 13, 1, January 1970.

Su, Stanley Y. W., Harper, K., "A Directed Random Paragraph Generator", Preprint No. 13, International Conference on Computational Linguistics, 1969.

Su, Stanley Y. W., "A Computational Model of Paragraph Production", Technical Report No. 71-102, Center for Informatics Research, University of Florida, 1971.

Wilks, Y., "Computable Semantic Derivations",SP-3017, System Development Corporation, Santa Monica, January 15, 1968.

2. A "Feed-Forward and Feed-Back" Controlled Operating System (with Truman C. Prewitt)

This research deals with the design of a "feed-forward and feed-back" controlled operating system configuration for a large scale time-sharing system. The project is divided into the following phases: a) an extensive data collection and evaluation of OS/360 to determine its performance characteristics, b) the design and implementation of a feed-back monitor which handles dynamically the allocation of system resources and c) the addition of a feed-forward component to provide the system with information useful for predicting future resource demands. The semantic information of users' jobs is utilized by the feed-forward component.

References

Coffman, E. G., "Analysis of Two Time-Sharing Algorithms Designed for Limited Swapping", J. ACM 15, 3, July 1968.

Denning, P. J., "Resource Allocation in Multiprocess Computer Systems", Technical Report MAC-TR-50, MIT, Cambridge, Mass., 1968.

Stevens, D. F., "System Evaluation on the Control Data 6600", IFIP International Conference 1968, pp. 34-38.

3. An Interactive Computer-Assisted Diagnosis System (with Sung-Yu Sheng)

This research deals with the design and implementation of a diagnosis system for assisting medical personnel to do proper diagnosis of a patient's ailment. The system consists of two major components: a question-answering module (QAM) and a diagnostic program module (DPM). The QAM gathers medical data directly from the patient through an on-line device and provides the data to the DPM to complete a diagnosis. Automated classification and hierarchy construction techniques are applied in the DPM to classify the diseases dealt with in the system. The DPM uses the resulting classification to assist medical personnel to formulate proper questionnaires which are programmed in the QAM to discriminate different diseases.

References

Greenes, Robert A., Barnett, G. Octo, Klein, W.W., Robbins, A., Prior, R.E., "Recording, Retrieval and Review of Medical Data by Physician - Computer Interaction", The New England Journal of Medicine, Vol. 282, No. 6, Feb. 5, 1970.

Kanner, Irving F., "The Programmed Physical Examination with or without a Computer", The Journal of the American Medical Association, Vol. 215, No. 8, February 22, 1971.

4. Automatic Speech Recognition Using Syntactic and Semantic Constraints (with A. Paige)

a. Introduction

There is an enormous range of potential applications for automatic speech recognition systems. Usually these applications involve the communications between man and machine in a form which is most convenient for man. More indirectly but nevertheless important they could also involve the improvement of speech communications between man and his fellow man. Unfortunately,the present state of art in speech recognition does not motivate great confidence that useful and reliable systems are within our immediate grasp. Since, however, the potential utility is so great, further investigation seems to be easily justified.

It would certainly be possible to give a very long list of potential applications ranging from communication aids for the deaf to automatic typewriters. In all probability, many of the possible uses have as yet to be enumerated. It is not very likely that a feasible solution will be available in the near future and therefore, speculations on potential applications are just that, speculations.

Perhaps the most realistic course to follow is to admit that a simple solution is not likely and that a more sophisticated approach based upon our knowledge of human perception of speech has the greatest potential for success. The system described here utilizes not only the details of the acoustic signal but also the syntactic and semantic restrictions which are inherent in any human speech utterance.

Simply stating what constitutes recognition is no small problem. However, for the present purposes automatic recognition may be thought of as a mechanical structure which responds in a way which is similar to the responses of a number of human observers to human produced speech material. This definition of course raises at least as many questions as it answers. The full range of human processing of speech is simply too large to hope to simulate by algorithmic process. However, it is reasonable to consider certain subproblems which are finite in scope and yet general enough to allow future expansion. For example, the language being considered for this system is a small subset of English. However, the description of this language is in the form of a grammar. Future inclusion of a larger part of English would then be accomplished by expanding the rules allowed in the grammar. Hence, this generalization is more a matter of degree of complexity than of principle.

The fact that syntactic and semantic constraints are being incorporated into the system proposed here is recognition of the failure of simple analysis of the acoustic signal by itself to function as an acceptable and useful speech recognizer. This is not surprising since contextual constraints seem to be a very important part of human perceptions of speech. In the language of communication engineers, syntax and semantics are being used for error detection and correction. The errors which are made are in the acoustic part of the recognition system.

The acoustic analyzer part of the system makes an initial classification of small segments of the utterance based largely on the details of the acoustic signal only. From the large number of such acoustic analyzers which have been constructed and reported in the speech literature, one can expect a rather high incidence of errors no matter how the device is constructed. There are two different courses to follow then. One would be to attempt to somehow do a better job analyzing the acoustic signal or alternately, as is proposed here, to use other information which is inherently present in the utterance but is not as readily available as are the acoustic aspects.

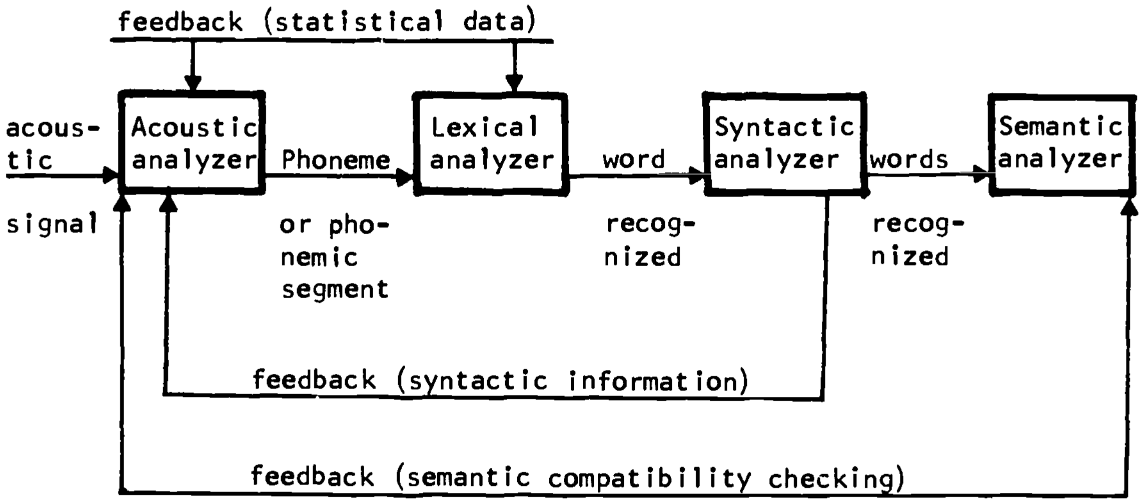
Most approaches to speech recognition to date have attacked the problem of segmenting the acoustic data of an utterance into units of phonemes. However, two problems encountered in the segmentation of the acoustic data are 1) the phonemic units segmented do not exist outside the context of the syllables and words in which they appear, and 2) the enormous number of possible ways to segment an acoustic sequence due to the errors made in the initial classification and the variations in human utterance. One solution to the first problem as proposed here is to segment an utterance into syllables or combination of syllables rather than individual phonemes. To solve the second problem mentioned above, it is proposed to make use of (1) the statistical information concerning the co-occurrence of phonemic segments, (2) the syntactic information concerning legitimate syntactic constructions and (3) the semantic information concerning word compatibility in an acoustic analyzer to reduce the number of alternatives at every stage of the segmentation of the acoustic sequence.

The next section of this proposal describes the model and its general operation as a speech recognition system. The proposed system will be implemented and experimented in parallel with the work on the classification of acoustic data which is discussed in a later section. The language to be used in the initial testing of the system is also described.

b. System Description

The basic approach in this system is to use the co-occurrence data of the syntactic and semantic information of the words already identified to optimize the initial classification of the subsequent acoustic data. The optimization will be achieved by the system's abilities to do error detection and correction, search path elimination and selection of probable path in the analysis based on statistical, syntactic and semantic properties of the segments which have already been recognized.

The following is a diagram of the operation of the proposed system.



Input to the acoustic analyzer is a sequence of acoustic signals which represents a human produced sentence. Spectral analysis will be performed alternatively on the beginning portion and the end portion of the input sequence to identify the syllables or combinations of syllables at both ends. The analysis will be performed on both ends because the beginning and the end of an utterance are the most reliably identified boundaries that the system can assume initially. Each identified segment will be passed to a lexical analyzer which provides information to the acoustic analyzer about the probabilistic measures for the possible phonemic segments following the current one. The probabilistic measures will be compiled on the basis of syllabic co-occurrences of the words in a preconstructed lexicon. The feedback information from the lexical analyzer will be used by the acoustic analyzer to reduce the number of considerations when the system analyzes the next segment. The different measures allow the acoustic analyzer to select the most probable syllables to match against the next input sequence.

When a word is recognized, the syntactic information stored in the lexicon entry will be retrieved and checked against the syntactic rules of the language (see section three). This is to determine the possible grammatical categories of the words which might follow the current one. These categories will be used by the acoustic analyzer to limit its considerations for the next segment among the segments of the words in those categories. The syntactic checking will take into consideration all the grammatical properties of all the words on both ends of the input sequence which have already been recognized by the system.

When more than one word is recognized, the semantic analyzer will be called upon to check their semantic compatibility by using the selection restrictions associated with the words. The semantic theory proposed by J. Katz (1964, 1967) will be adopted for this purpose.

Any grammatical incompatibility or error found by the syntactic analyzer and the semantic analyzer will cause the acoustic component to back-track to a point where it can make an alternate pass in its analysis. This "backtracking" operation will be aided by the system's recording of the well-formed segments recognized in the previous paths of analysis.

(1) The Lexicon

The lexicon to be constructed for the proposed system is a listing of syllables. It may also contain some combinations of syllables which are to be recognized as units. Each entry in the lexicon will contain some statistical information concerning which syllables are likely to follow the entry and also which are likely to precede the entry (the analysis of the input acoustic sequence is done on both ends). The statistical information will be compiled on the basis of the frequency of co-occurrence of the syllables in a preconstruct word list for the system. The entries will be associatively linked together to mark the words that can be formed by the entries. Associated with each word will be syntactic markers to indicate its syntactic properties, semantic features to represent its semantic properties and also semantic restrictions to specify the context in which the word can occur. The data in the lexicon will be used by the various components of the system as described in the preceding section.

(2) The Input Language

For the initial implementation of the system, we will use an English-like language whose syntactic rules and vocabulary are much more restricted than natural English. For example, the data manipulation language used in many information systems to process and retrieve data stored in a large data base. The following set of statements illustrates the type of language being considered. It is a modification of the data manipulation language used in IBM's System/360 Generalized Information System.

```
QUERY PERSONNEL
LOCATE SALE PERFORMANCE RECORD
WHEN PRODUCT NUMBER EQUAL 1723
AND MONTH EQUAL 6803
AND NET GREATER OR EQUAL TO 1000
LOCATE PERSONNEL RECORDS
LIST NAME, LAST DEPARTMENT NUMBER, LAST RATE
EXHAUST PERSONNEL
END PROCEDURE
```

The reasons for choosing this type of language for the initial model are twofold. First, the language is simple enough and is thus suitable for initial implementation and testing of the algorithm and technique to be used in the system. The proposed system will be built to allow future relaxation of the input language constraints. For example, the system will accept the grammatical rules as data for doing syntactic checking. Future modification of the input language amounts to changing the grammatical rules without requiring the reprogramming of the syntactic analyzer. Similarly, the semantic features and selection restrictions will be used by the semantic analyzer to determine the compatibility of word-combination. If the input language is changed, only lexical entries need be modified but not the semantic analyzer itself. Secondly, the success of recognizing this type of language would find immediate applications in the field of information science. This proposed system will allow the user of an information system to talk to a machine and command the machine to retrieve data from an information store.

The system proposed in this portion will be studied and implemented independently of the component which carries out the spectral analysis and the classification of acoustic data as described later. It will be assumed initially that syllables or combinations of syllables will be made available by the acoustic analyzer. With controlled perturbations on the syllables which simulate the situation that voices existing in the original acoustic sequence or that errors were made in the initial classification, the acoustic analyzer will carry out the error checking and correction and also the prediction of the next syllable on the basis of the syntactic and semantic information available up to the point of analysis.

The first task in this research will be to construct the lexicon described. The proposed components will then be programmed to test the proposed general operation of the system.

c. General Description: Acoustic Analyzer

The purpose of the acoustic part of this recognition system is to supply a preliminary classification of the speech utterance and it is on the results of this portion of the system that the syntactic/semantic part operates. The present results of the syntactic and semantic parts will also be available to the acoustic analyzer to aid in this first order classification.

Another section of this report details an investigation of the problem of transforming the acoustic signal representation of a speech utterance into an articulatory configuration. The purpose of this transformation would ultimately be to classify the speech utterance and would be used in the acoustic part of the overall recognitions system described here. Unfortunately, a large amount of work remains to be done on this transformation before it becomes useful. Therefore other approaches which hold promise of simpler implementation must be considered for use in the acoustic recognition section.

The acoustic recognizer will operate on units of speech of approximately syllabic length. The details of a study of these units are given in another section. However, briefly, the "machine" syllable will be determined as that piece of the input speech signal which lies between two major minimas in the amplitude of the signal. The main characteristic of such a segment which is of interest to this discussion is that it is small, approximately the same length as "perceptual" syllables. Presuming that there is a strong (although not necessarily 1:1) correspondence between the perceptual and machine syllables it is then possible to say that there are a reasonably small number of such units in a vocabulary of substantial (i.e., useful) size. It is, of course, highly desirable that the units on which the acoustic recognizer work be small in order to minimize both storage of the units and search time of them. On the other hand, recognition accuracy is expected to increase with the size of the unit being examined. Syllables then seem to be useful to construct a reasonable sized vocabulary and yet are small enough in number so that they can be search (exhaustively if necessary) in a modern digital computer.

The storage of speech utterances of syllabic length will be accomplished by a method described by Paige (1970). Essentially this method stores a compact spectra representation of the syllable, in much the same sense that a single set of numbers (the spectrum) describes a steady state vowel. In the case of the syllable a few sets of numbers (on the order of three to five) may be used to describe the syllable. This generalized spectrum has yet to be investigated very thoroughly however, especially as to its ability to reliably differentiate between similar utterances.

The general procedure for obtaining the compact spectral description discussed above is not dependent upon the functions with respect to which the spectrum is computed. Usually the spectrum of speech signals is referred to sinusoidal functions and these have proved to be very useful especially in describing the acoustic characteristics of speech. For the speech recognition system described in this section spectral analysis is to be done by digital computation. If there is a large amount of speech to be processed the spectral computations can become very time consuming and expensive. For this reason an alternative set of basis functions is being investigated.

Computation of the spectrum of a signal with respect to sinusoidal functions requires a large number of multiply and add operations. On the other hand, the computation of the Hadamard transform requires only (the same number of) add operations. Hence, because there are no multiply operations, it should be cheaper and faster to compute the Hadamard transform than the Fourier transform. The recent introduction of the fast Fourier Transform substantially reduces the amount of time necessary to compute the Fourier spectrum. However, a fast Hadamard transform also exists and results in the same order to magnitude reduction.

While the Hadamard transform can be more rapidly computed than the Fourier transform the question arises as to its utility for describing speech signals. The nature of the Hadamard functions indicates that the transform will be used for detecting zero crossing rate, which is the kind of information which is useful in describing speech signals. Some preliminary computations of the Hadamard transform of analytical signals which are of interest to speech work have been carried out. Qualitatively, the results are similar to ordinary Fourier spectra. Comparison of Hadamard and Fourier spectra for human produced speech must now be carried out.

d. Sample 1 - The Hadamard Transform of Speech Signals

As noted in the previous section the principle justification of the Hadamard transform is its simplicity. The utility of this transform for dealing with speech signals must be determined, especially in the overall context of the speech recognition system under investigation. This evaluation is to proceed in two phases. The first is described in this section and the second is described in the following section which deals with compact spectral descriptions of short dynamic utterances (syllables).

In the first phase the Hadamard and Fourier spectra are to be compared directly with each other for simple human produced utterances. It will be understood that Fourier spectra will be determined both by digital computation and sonograph analysis.

Several different experiments are proposed and it is understood that succeeding ones will be carried out only if results to that time indicate that the Hadamard transform does possess useful properties for speech signal analysis. Otherwise, the investigation will be discontinued.

In addition to the spectra itself, the Hadamard transform may also be used to determine periodicities of complex signals such as the fundamental frequency (or pitch) associated with speech signals. This periodicity manifests itself in the Hadamard spectrum in the same way that it appears in the Fourier spectrum.

The Cepstrum has proved to be a very useful method of measuring this fundamental frequency and it is proposed that the analogous operations be performed with the Hadamard spectrum. This also may prove to be a simple means of detecting the presence or absence of voicing in the speech.

Specifically, the Cepstrum of sentence length utterances will be computed and the fundamental frequency contour throughout the utterance will be determined and compared with that obtained using Fourier analysis. There should be little doubt at the completion of this experiment as to the usefulness of the Hadamard transform as a pitch detector.

Specific utterances used will be chosen from the vocabulary that is initially chosen for use with this recognition system. A variety of male and female speakers will be used to determine that the proposed system operates satisfactorily on a wide range of fundamental frequencies.

The evaluation of the Hadamard transform for the spectral representation of individual speech sounds will begin by comparing Hadamard and Fourier spectra directly for a number of vowels produced in isolation by several speakers, both male and female. Specifically, the formants of the vowels will be determined by examining the various spectra. The formant frequencies are to be determined by several human judges and then the results are compared. Several variables in the computation of the spectra must also be investigated; specifically the length of the data whose spectrum is to be determined (length of impulse response of filter) and the shape of the window used to look at the data. This is a reasonably significant point since preliminary investigations have shown that these are rather large sidelobes in the Hadamard filters. This phenomenon occurs in Fourier spectra as well, and in this case properly shaped windows have proved useful in "smoothing" the frequency response.

The next series of experiments deals with more realistic samples of speech; dynamic utterances of syllabic length. These types of utterances are of direct concern to the overall speech recognition system since they are the basic units which are classified in the acoustic part of the system. The principle concern will be with determining how frequently the spectrum must be computed and over how long a period successive spectra are to be averaged. The success of the compact spectral representation described in the following section is dependent upon relatively small variations in the spectra over the duration of the syllable. The more variation there is, the less compact is the representation.

If the spectra must be computed frequently (say every sample) the computation time increases greatly. The possibility that frequent computaion must be done arises from the fact that the Hadamard power spectrum is not phase invariant to the sum of a number of sinusoids. It may then be necessary to compute spectra frequently and average the results. If the averaging time becomes comparable with significant changes in formant frequencies, the results could have unacceptable time resolution.

Inspection of the results of the Hadamard transform as a function of time will yield only qualitative estimates of its utility. More realistic tests will result when the compact spectral representation as determined in the following section is investigated.

e. Sample II - Compact Spectral Representation

The simplest way to represent a dynamic utterance in a computer would be to store successive spectral vectors. This,unfortunately, would use up a large amount of storage. Furthermore,comparing the stored utterance with some utterance which is to be classified could prove to be very time consuming and inaccurate.

Paige (1970) suggests a method for representing the spectral part of a dynamic utterance in a computer in a very compact form. This procedure has been tested only on a limited number of syllables and it is proposed to test this on a larger variety of data and also to test it using the Hadamard spectra instead of the Fourier spectra.

An initial experiment which by itself may be a useful although very restricted recognition system will limit the vocabulary to the ten digits. Some initial experiments with such a limited vocabulary have indicated that reasonably high recognition scores can be obtained. Several problems however, arise. One concerns normalization for the speaker. That is the spectral representation for a particular syllable is different for different speakers just as are his spectral representations isolated vowels. An initial normalization of the spectrum will be attempted using a single frequency transformation of the form $f \rightarrow kf$ where k is determined by an initial training procedure carried out on vowels. A second problem which must be investigated concerns the actual metric which is to be used in comparing the unknown utterance with a library of standards. There are several possibilities for doing this and each will be examined.

The next study will be similar in nature to the one described above but will involve a much more expanded vocabulary of syllables which will be chosen from the system vocabulary. One of the purposes of the study will be to supply a more realistic estimate of the kind of errors which the syntactic and semantic parts of the system must detect and correct. In addition, of course, it will be used to test in a more comprehensive way the results of the previous experiments which used the very limited vocabulary of ten digits.

f. Significance of the Proposed Research

The use of electronic devices as speech recognizers has revealed great variations in the acoustic properties of human utterances even in the utterance of the same word or sentence. The variations in the speech of different individuals has greatly complicated the problem of building a speech recognition system. For example, for a given speech sound it is very difficult to determine which phoneme or phonemic segment has these acoustic properties. In general, there will be many phonemic segments whose acoustic descriptions are similar to the speech sound because of the great variations associated with each of them. To program a system to test for all the possibilities is too time-consuming and costly. A method must be employed to limit the number of tests of the possible phonemic segments, and to select the most probable one for testing first.

The proposed recognition system intends to make use of statistical data of syllable co-occurrence, and syntactic and semantic information of the recognized words to reduce the number of considerations in the component which does the spectral analysis and classification of acoustic data. It will speed up the process of analyzing acoustic data and enable accurate recognition of human speech.

It is well known that an automatic speech recognition system has many potential applications. The achievement of the initial goal of recognizing the sentences of a data manipulation language would have great practical applications in the field of electronic data processing. A speech recognition component would replace a console typewriter and allow a machine operator to give commands to the computer directly in spoken form. A more important application is that computer users can then bypass the undesirable punched card medium and simply tell the computer what needs to be done.

References

Katz, J. J. and Foder, J. A., The structure of a semantic theory, in The Structure of Language, J. A. Foder and J. J. Katz (eds.), Englewood Cliffs, New Jersey: Prentice-Hall, 1964.

Katz, J. J., The Philosophy of Language, New York: Harper and Row, 1966.

Paige, A., A spectral characterization of short dynamic utterances, Communication Sciences Laboratory Quarterly Progress Report, 8.1, University of Florida, March, 1970.

Su, S. Y. W., Managing semantic data in an associative net, In the Proceedings of the ACM Symposium on Information Storage and Retrieval, University of Maryland, 1971.

William A. Yost

I. Personal Research

1. Studies in Psychoacoustics

The following proposals constitute a research program in Psychoacoustics. Proposal A is concerned with certain interactions of two auditory cues, time and intensity, which are responsible for localization. The research and ideas represent a continuation of current research being conducted by the applicant. Thus, it is narrow in some respects, but the implications for other areas of binaural hearing suggests that the project has a wide spectrum of applications as its goal. Proposal B is concerned with an area of research that has not been studied in any depth by psychoacousticians: that of temporal or phase discrimination called Auditory Temporal Acuity. This is a new area of study since most research in psychoacoustics deals with frequency and intensity discriminations and newness of the field makes it a challenging area of research. However, within the field it is somewhat difficult to predict which lines of research would be most fruitful. In this regard, the applicant has indicated one set of parameters which requires further investigation.

a. Interaction of the cues responsible for localization

(1) Theoretical Framework

Although there is a long history of the study of binaural hearing, many interesting problems remain unanswered. The early physical work of Wightman and Firestone (1930) and the psychophysical studies of Stevens and Newman (1934) showed that two cues, interaural time (phase) and interaural intensity, were required to locate an acoustic stimuli, but that the interaural temporal differences were most easily measured for low frequency stimuli. Stevens and Newman argued that interaural intensity was the important cue for localizing high frequency stimuli and that interaural time was critical for locating low frequency stimuli. The question of the interaction of these two cues has been left relatively unanswered since Stevens and Newman first proposed the Duplex Theory of localization. That is, how do the physical differences in time and intensity present in the waveforms of acoustic energy arriving at the ears combine to enable the observer to localize stimuli of different spectral content.

For several reasons the free field localization situation is not a convenient condition to test this general question. First, the measurements of the two differences are difficult to make. Second, the two differences covary for a free field sound source so that independent control of the differences is impossible. These problems can be avoided by presenting either a given interaural temporal difference, a given interaural intensive difference, or both, to an observer via headphones. With headphones the "image" appears to lie within the head on a line between the ears; hence, the locating task is called lateralization. As in the free field, the lateralized image appears toward the ear leading in time, or having the greater amplitude. Thus, a lateralization task is often used to study the role of the two cues of localization, time and intensity.

Although there is a rich history of studies in lateralization, few studies have investigated directly the interaction between the interaural intensive difference and the interaural temporal difference. These studies have presented the two differences in opposition: the ear leading in time receives the stimulus with smaller amplitude (Deatherage, 1966). Some recent investigations (Hafter and Jeffress, 1968) suggest caution in the interpretation of results from such studies since observers may hear two images: one assoicated with the temporal difference, the other associated with the intensive difference. Virtually no lateralization studies exist in which the two differences agree. Thus, one goal of this proposal is to study the interactions between the interaural temporal and the interaural intensive differences (both differences agree) required to lateralize stimuli of various frequency contents.

Data obtained in lateralization experiments involving only the interaural temporal difference and stimuli of different frequency content suggest the second goal of this proposal. As early as 1940 Huges demonstrated that for pure tones interaural phase rather than interaural time determined an image's location: the lateralized image appears toward the ear leading in phase for a phase difference of less than 180°, but toward the ear lagging in phase for a phase difference greater than 180°. Many investigators (Zwislocki and Feldman, 1956; Klumpp and Eady, 1956; Yost, 1971) have reported that a constant phase difference , rather than a constant temporal difference, is required for discrimination between the location of two lateral images. When complex stimuli are used, however, other investigators (Teas, 1962; Deatherage, 1966; Yost, et al, 1971) have shown that interaural time, not phase, is the determining factor for such lateral discriminations. Thus, another goal of the proposal is to study the interaction of the interaural temporal-phase difference in lateralization. The complexity of a stimulus is varied.

The two goals of the proposal are not only relevant to the problems stated above, but also basic to many other unanswered questions involving binaural hearing:

(a) The time-phase paradox in localization

Stevens and Newman showed that the location of a sound source appears to depend on the interaural temporal difference, in that two sound sources in the same location appear in this location even if their frequencies are disparate. Since time of travel from source to ear is independent of frequency, the two sources produce the same temporal difference. Mills (1958), however, measured the phase and temporal differences present when an observer is asked to discriminate between the two sources at different locations. He found that a constant phase difference rather than a constant temporal difference best described the discrimination when the frequency of the stimuli was varied. These two sets of findings indicate a "position" paradox. For instance, a 200 Hz and a 500 Hz tone will be localized at the same position if their sources are placed together. An observer will require a constant phase change for both tones to be moved the same "perceived" distance to a new location. Once moved, they will actually be in different positions since a larger temporal change (larger move) will be required for the 200 Hz tone than for the 500 Hz tone. Of course, once they have been moved by the same phase differences, the observers should perceive them not in the same but in different places. Thus, an equivalence in one sense produces inconsistency in another, i.e., a paradox.

One possible solution to this paradox involves a better understanding of the interaction between the intensive difference and the phase-temporal difference. The intensive differences is usually assumed to be negligible at low frequencies, but recent measurements in lateralization (Elfner and Perrott, 1967) indicate that the intensive differences available at low frequencies in the free field should not be considered negligible. Thus, the intensive differences could be contributing significantly in determining the location of a low frequency stimulus.

(b) Externalization of the lateral image

As was mentioned earlier, when stimuli are presented to listeners via headphones, the image is perceived within the head. Toole (1970) showed that most attempts to make the lateral image appear external to the observer have failed. Daumaske (1971) reported that stereo listening in a free field can be greatly improved if the recording and playback systems take into account the "cross-talk" occurring between the ears in a natural listening situation. This observation suggests that a lateralized image could perhaps be externalized if the proper "cross-talk" were introduced artificially via the acoustic apparatus. That is, the "cross-talk" between waveforms arriving at the ears may be responsible for an externally-located source actually sounding external. If so, then "externality" can be synthesized via earphone simulation. The "cross-talk" to be introduced is based on estimates of the intensive and phase-time changes which occur as sound moves around the head. Therefore, before one can introduce "cross-talk", he must decide whether it is interaural phase or time which is the important variable in lateralization. And more importantly, one should measure how the phase-temporal differences interact as a function of changing the stimulus from a pure tone to a complex stimulus. Once the phase-time interactions are better understood, the "cross-talk" hypothesis for externalizing a lateral image may be studied.

(c) The relationship between binaural masking and lateralization

A large body of literature has been devoted to the study of the binaural masking level difference (MLD, see Green and Henning, 1969). The MLD is the difference between the signal-to-masker ratio required for detection in a dichotic listening condition and that required in a dichotic condition. One example of a dichotic condition is a situation in which a wide-band Gaussian noise is presented as a masker identically at both ears and a pure tone signal is added to the masker at one ear 180^{o} out of phase relative to the other ear (^{S}180 case). One diotic condition consists of the signal and masker being presented identically at both ears (So case). In this case, the signal-to-masker ratio required for detection is usually 15 dB smaller in the ^{S}180 case than in the So condition; the MLD is thus 15 dB. For the dichotic condition, but not the diotic condition, both an interaural temporal and an interaural intensive difference exist. Many investigators (Hafter, et al, 1969; Jeffress, et al, 1956) have argued that it is these interaural differences which lead to the MLD. One problem in studying the role of the interaural differences and the MLD, however, is that since the masker is a random noise, these interaural differences are random variables. Thus, complicated statistical assumptions are required to even measure the interaural differences.

One attempt to circumvent this problem is to use a pure tone masker (not a random masker) so that one can measure the interaural differences and the MLD in a tone-on-tone binaural masking study (Hafter, et al, 1970; Robinson and Yost, 1970). In the pure tone masking situation, the interaural phase difference appears to be the more important variable for determining the MLD. But no one yet has ascertained how the phase or temporal differences vary as the masker is made more complex (i.e. a random noise). Therefore, we will study the interactions between the phase-time difference in lateralization for pure and complex stimuli to shed some light on the interaction between the interaural differences and the MLD. In addition, as was described above, both intensive and temporal differences exist for the dichotic condition. A better understanding of the interaction between these two differences in a lateralization procedure might help to better understand the relationship between the MLD and the interaural differences.

(d) Fused images for high frequency complex stimuli

Zwislocki and Fledman (1956) showed that for pure tones greater than approximately 1500 Hz the lateral image could not be moved by the introduction of any interaural temporal difference. In fact, high frequency tones appear not to have a well-defined lateral position. However, broad-band stimuli with components only above 5000 Hz do have a well-defined location and may be moved by introducing a temporal difference, although the required temporal difference is several times larger than that required for low frequency broad-band stimuli (Yost, et al, 1971). Again there is a discrepancy between simple and complex stimuli and the role of the interaural temporal-phase difference in lateralization.

(e) The acuity of the binaural system

The interaural temporal difference required to discriminate a change in location of a lateral image can be as small as 10 microsecs. A perplexing physiological question concerns this microsec resolution, given a neurological system the elements of which operate with a 100 millisec resolution. A better understanding of the interactions among interaural phase, interaural time, and frequency might help explain the good temporal acuity of the binaural system.

In summary, then, it is the interaction between the stimulus parameters of interaural time and interaural intensity investigated for simple, i.e., sinusoid and complex, i.e., random noise acoustic signals which seem to underlie many questions concerning the binaural system's ability to localize a sound source. By studying the interaction of the intensive and temporal differences in a lateralization paradigm and the temporal-phase difference for simple and complex stimuli, many fundamental questions about the basic functions of hearing may be answered.

(2) Methods

In general, a lateralization paradigm will be used in which observers will be asked to either discriminate between the location of two images or match the location of one image to that of another image. Human observers will be hired on an hourly basis and tested in the Psychoacoustics Laboratory of the Communication Sciences Laboratory. The discrimination tasks will be either a two-alternative temporal, forced choice procedure (2AFTC) or a SAME-DIFFERENT procedure. The interaural temporal and intensive differences will be monitored both electrically and acoustically. The overall level of stimulation will be kept low enough (less than 70 dB SPL) to avoid harmful stimulation.

(3) Detailed Description of the Experiments

As a first step in understanding the intensive-temporal interactions, a discrimination experiment involving both interaural differences will be conducted. In a 2AFTC lateralization task observers will be asked to discriminate between two stimuli differing only in interaural time (T and T $+\Delta$ T) or only in interaural intensity (I + I $+\Delta$I). Once threshold values of Δ T and Δ I have been obtained, then discriminations of the nature, T vs. T $+\Delta$ I and I vs. I + Δ T will be conducted. Next, discriminations of the nature T $\pm\Delta$ I vs. I $\pm\Delta$ T will be performed. In all these experiments the ear leading in time will also be the ear with the greater amplitude. Also, discriminations involving the intensive difference (I, Δ I) will have to be counterbalanced such that any cues associated with a change in loudness rather than a change in lateral position are randomized. Finally, these discriminations should be investigated at different frequencies. This set of experiments should indicate how the two cues combine to produce just-detectable changes in the location of a lateral image. Hopefully, a rule or model will be developed to describe the interaction of the two interaural differences. (One such model has already been proposed by Hafter and Carrier (1970) and by Yost (1971), but it has not yet been thoroughly tested.)

As a first step in understanding the phase-time problem for lateralization, a two-tone complex stimulus will be studied. For this experiment two frequencies of equal amplitude and duration (both frequencies gated with a 25 msec rise-decay time) will be played to the two ears: F and F $+\Delta$ F to the right ear and F and F + Δ F to the left ear. Let us call stimu.us F the standard frequency, and F $+\Delta$ F the comparison. The experimenter will alter the lateral location of the standard frequency by varying its interaural temporal difference, and the observer will be asked to vary the temporal difference of the comparison frequency until the image is perceived to be in the same place (location) as the standard. For large values of ΔF and for frequencies of less than 1500 Hz, the observer will probably match the two stimuli when the two frequencies have the same interaural phase difference. However, as ΔF decreases and/or one or both freuqencies becomes larger than 1500 Hz, the relation between the standard and comparison interaural delays will become unclear. Of course, it is the purpose of this study to determine these interactions. Since the experimenter has no way of judging if the observer has made a "a correct match," many trials will be presented for each condition, and a statistical analysis of the data will be given careful consideration. In addition, some of the conditions will be used in a SAME-DIFFERNT discrimination paradigm. That is, the observer will be presented two sets of two-tone complexes. Dn the SAME trials the two, two-tone complexes will be chosen as in the original matching experiment in which they were judged to be in the same lateral location. Dn the DIFFERENT trials the two, two-tone complexes will be in different places. Thus, one may plot some measure of percent correct discrimination versus either interaural time or Δ F, depending on which parameter is used to separate the position of the two, two-tone complexes. This procedure requires that all other cues except position (i.e., frequency) be randomized. The SAME-DIFFERENT procedure allows one to study the interaction mentioned above in a more deterministic manner than does the matching procedure. However, the SAME-DIFFERENT task is a more time-consuming psychophysical procedure.

(4) Summary

Two sets of interactions involving the interaural differences required to localize a sound source have not previously been studied in a detailed manner. The interaction between the temporal and intensive differences required for lateralization and that between the phase-temporal differences required for lateralizing simple and complex stimuli appears to be important for the understanding of a variety of problems involving binaural hearing. Dnce the initial experiments described in Section III have been conducted, the research could take many directions. It could proceed deeper into the basic questions involving these interactions, or research might be undertaken in one or more of the related areas mentioned in Section I. Regardless of which experiments are conducted after the two described in Section III, the results from these two should provide useful information concerning the two sets of interactions which underlie a broad spectrum of problems in binaural hearing, an active and productive psychoacoustics program will result from pursuing the goals of this proposal.

References

Daumaske, Appear in J. Acous. Soc. Amer., 1971.

Deatherage, Bruce H., Examination of binaural interaction, J. Acous. Soc. Amer., 39, 232-242, 1966.

Elfner, L. and Perrott, D., Lateralization and intensity discrimination, J. Acous. Soc. Amer., 42, 441-445, 1967.

Green, D. M. and Henning, G. B., Audition, in Annual Review of Psychology, 2D, 1D5-128, 1969.

Hafter, E. R., Burbon, W. T., Blocker, A. S. and Tucker, A., Direct comparison between lateralization and detection under conditions of antiphasic masking, J. Acous. Soc. Amer., 46, 1452-57, 1969.

Hafter, E. R. and Carrier, S. C., Masking level differences obtained with pulsed tonal masker, J. Acous. Soc. Amer., 46, 14D1-1D, 1969.

Hughes, J. W., The monaural threshold: Effect of subliminal contralateral stimulation, Proc. Roy. Soc., 486-490, 1938.

Jeffress, L. A., Blodgett, H. C., Sandel, T. T. and Wood, C. L., Masking of tonal signals, J. Acous. Soc. Amer., 28, 416-426, 1956.

Klumpp, R. G. and Eady, E. R., Some measurements of interaural time difference thresholds, J. Acous. Soc. Amer., 859-860, 1956.

Mills, A. W., On the minimum audible angle, J. Acous. Soc. Amer., 30, 237-246, 1958.

Robinson, D. E. and Yost, W. A., Lateralizability and detectability of a low frequency tone as a function of signal plus masker phase, J. Acous. Soc. Amer., 45, 336 (A), 1969.

Stevens, S. S. and Newman, E. B., The localization of pure tones, Proc. Nat. Acad. Sci., 20, 593-596, 1934.

Teas, Donald C., Lateralization of acoustic transients, J. Acous. Soc. Amer., 34, 1460-65, 1962.

Toole, F. E., In-head localization of acoustic images, J. Acous. Soc. Amer., 48, 943-949, 1970.

Wightman, E. R. and Firestone, F. A., J. Acous. Soc. Amer., 2, 271-280, 1930.

Yost, W. A., Tone-on-tone masking for the Mo-So, Mo-S and Mo-Sm conditions, J. Acous. Soc. Amer., 49, 103 (A), 1971.

Yost, W. A., Interaural delay discrimination, paper presented at Eighty-first Meeting of the Acoustical Society of America, 1971.

Yost, W. A., Wightman, F. L. and Green, D. M., Lateralization of filtered transients, paper presented at Eighty-first Meeting of the Acoustical Society of America, 1971.

Zwislocki, J. and Feldman, R. S., Just noticeable differences in dichotic phase, J. Acous. Soc. Amer., 28, 860-864, 1956.